Zuzanna Słobodzian
nr albumu: 412204

Kraków, 10.06.2024

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Sprawozdanie z projektu zaliczeniowego Data Mining: dane CO2 Emissions

1. Wstęp

Celem projektu jest wykonanie dwóch modeli regresyjnych na danych dotyczących zmiany wskaźnika emisji CO2 w różnych państwach świata na przestrzeni lat 1960–2014. Wybrane modele to kNN regresyjny oraz sieci neuronowe MLP, modele przewidują wartość wskaźnika w roku 2014, jest to ostatni rok, dla którego są dostępne dane. Projekt został wykonany w języku R, przy użyciu bibliotek: ggplot2, caret, neuralnet.

2. Opis danych

Oryginalnie dane mieszczą się w dwóch arkuszach, pierwszy zawiera nazwy i kody państw, nazwę i kod wskaźnika oraz wartości wskaźnika z lat 1960–2014. Drugi zawiera nazwy i kody państw, informacje o grupie dochodowej i regionie oraz dodatkowe uwagi.

Przed przystąpieniem do analizy, informacje z obu tabelek zostały połączone w jedną, tak aby uzyskać komplet informacji o nazwie państwa, przypisanym regionie, grupie dochodowej i wysokości wskaźnika na przestrzeni lat (rys. 2.1).

	A	B	C	D	E	F	G	H	I	J	K
4											
5	Country Name	Country Code	Indicator Name	Indicator Code	Region	IncomeGroup	1960	1961	1962	1963	1964
6	Aruba	ABW	CO2 emissions (metr	EN.ATM.CO2E.PC	Latin Amer	High income					
7	Afghanistan	AFG	CO2 emissions (metr	EN.ATM.CO2E.PC	South Asia	Low income	0.04606	0.053604	0.073765	0.074233	0.086292
8	Angola	AGO	CO2 emissions (metr	EN.ATM.CO2E.PC	Sub-Sahar	Lower middle income	0.097472	0.079038	0.201289	0.192535	0.201003

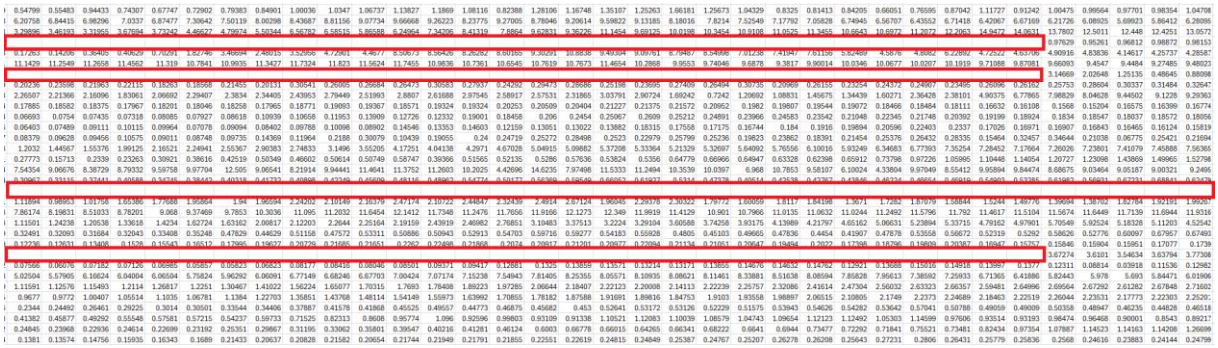
rys. 2.1

Kolumna „Country Name” zawiera nazwy państw oraz zbiorczych regionów jak np. *„East Asia & Pacific (excluding high income)”*, a kolumna „Country Code” trzy literowe unikalne id dla państw i zbiorczych regionów. Wartości „Indicator Name” i „Indicator Code” dla każdego wiersza są takie same: *„CO2 emissions (metric tons per capita)”* oraz *„EN.ATM.CO2E.PC”* – cały arkusz danych zawiera informacje o tylko jednym wskaźniku. W kolumnie „Region” jest wyróżnionych siedem grup: *„South Asia”, „Sub-Saharan Africa”, „Europe & Central Asia”, „Middle East & North Africa”, „Latin America & Caribbean”, „East Asia & Pacific”, „North America”*. Kolumna „Group Income” dzieli dane na cztery grupy: *„Low income”, „Lower middle income”, „Upper middle income”, „High income”*. Wartości w kolumnach zatytułowanych numerami kolejnych lat od 1960 do 2014 zawierają dane numeryczne oznaczające wysokość wskaźnika emisji CO2, liczonego jako tona CO2 na osobę.

3. Preprocessing i analiza danych

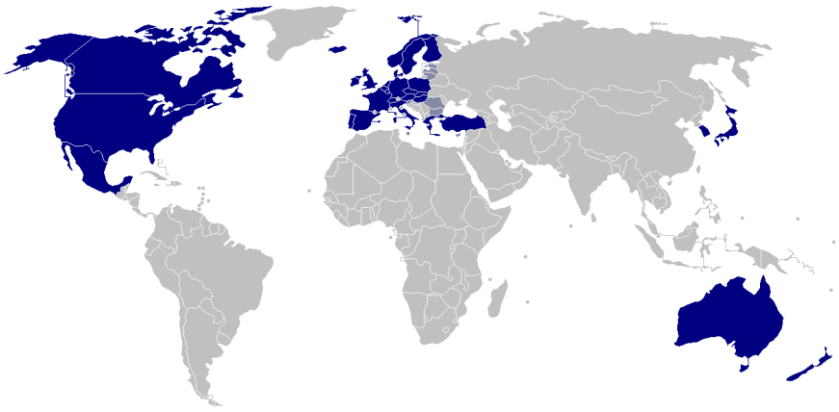
Po wczytaniu danych usunięto trzy górne puste wiersze, nadano odpowiednie nagłówki i usunięto wiersze z brakami danych.

Zdecydowano się na usunięcie całych wierszy zawierających braki danych, ponieważ układ występowania wartości NA uniemożliwiał zastąpienie ich konkretnymi wartościami. W przypadku kolumn z wartościami wskaźnika były to albo wiersze całkowicie puste, albo takie w których dłuższe ciągi braków wartości występowały obok siebie, zwykle w początkowych latach zbierania danych (rys. 3.1).,

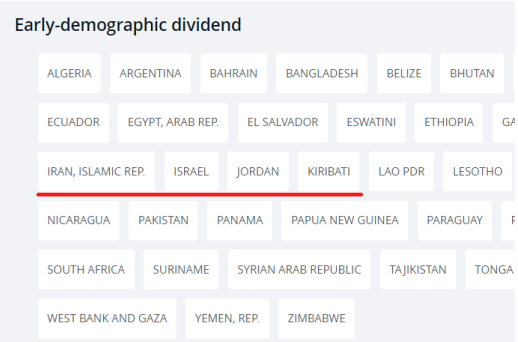


rys. 3.1

W sytuacji gdzie braki danych występowały w kolumnach „Region" i „IncomeGroup" dotyczyły one zbiorczych regionów (np. *„OECD members"*, *„South Asia (IDA & IBRD")*, *„Least developed countries: UN classification"*). Jako że zbiorcze regiony zawierają państwa położone w różnych częściach globu (*„OECD members"* (rys. 3.2)) lub należące do różnych grup dochodowych (*„Early-demographic dividend"* (rys. 3.3, rys 3.4)) nie było możliwe przypisane im wartości w tych kolumnach.



rys. 3.2



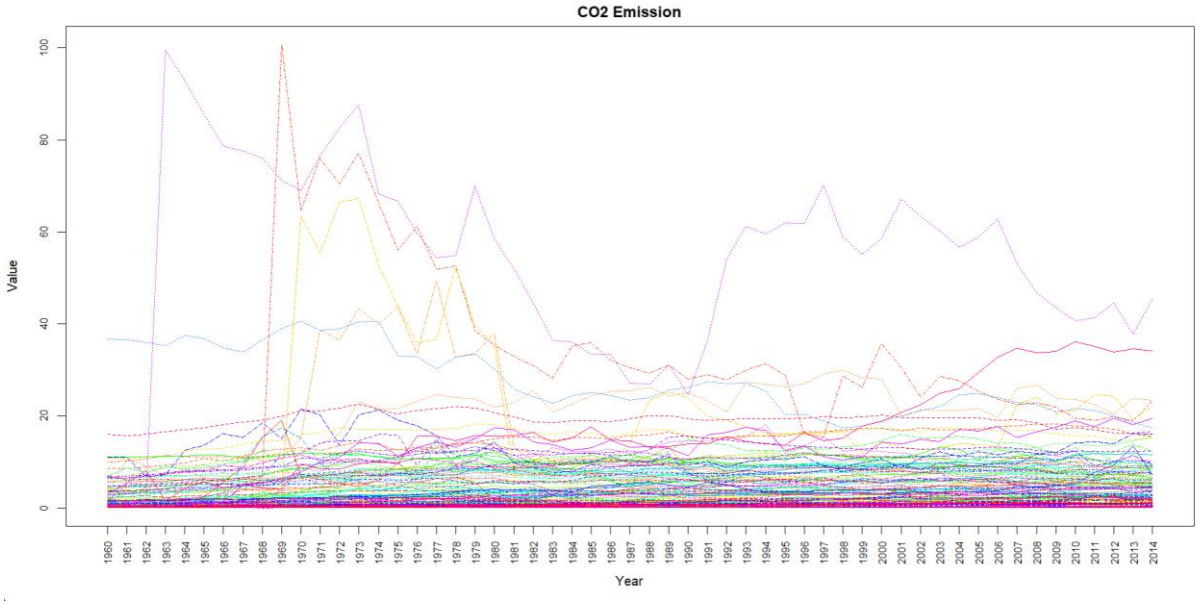
rys. 3.3

Ireland	IRL	CO2 emiss	EN.ATM.C	Europe & C	High income
Iran, Islamic Rep.	IRN	CO2 emiss	EN.ATM.C	Middle Eas	Upper middle income
Iraq	IRQ	CO2 emiss	EN.ATM.C	Middle Eas	Upper middle income
Iceland	ISL	CO2 emiss	EN.ATM.C	Europe & C	High income
Israel	ISR	CO2 emiss	EN.ATM.C	Middle Eas	High income
Italy	ITA	CO2 emiss	EN.ATM.C	Europe & C	High income
Jamaica	JAM	CO2 emiss	EN.ATM.C	Latin Amer	Upper middle income
Jordan	JOR	CO2 emiss	EN.ATM.C	Middle Eas	Lower middle income
Japan	JPN	CO2 emiss	EN.ATM.C	East Asia &	High income
Kazakhstan	KAZ	CO2 emiss	EN.ATM.C	Europe & C	Upper middle income
Kenya	KEN	CO2 emiss	EN.ATM.C	Sub-Sahar	Lower middle income
Kyrgyz Republic	KGZ	CO2 emiss	EN.ATM.C	Europe & C	Lower middle income
Cambodia	KHM	CO2 emiss	EN.ATM.C	East Asia &	Lower middle income
Kiribati	KIR	CO2 emiss	EN.ATM.C	East Asia &	Lower middle income
St. Kitts and Nevis	KNA	CO2 emiss	EN.ATM.C	Latin Amer	High income

rys. 3.4

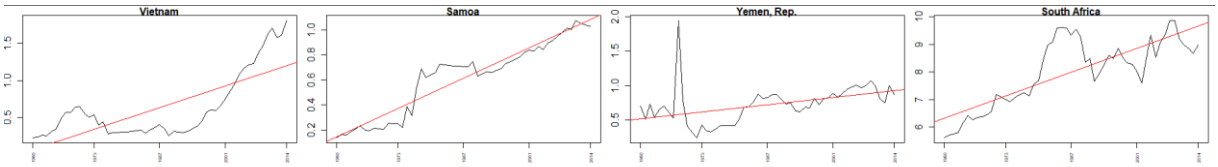
Kolumny „Indicator Name” i „Indicator Code” zostały usunięte, ponieważ nie przynoszą one żadnych informacji o danych – ich wartości dla każdego wiersza są takie same.

Następnie wykonano wykres porównawczy dla zmian wysokości emisji CO2 na przestrzeni lat dla wszystkich państw. Można zauważyć cztery linie: fioletową, czerwoną, żółtą i niebieską wybijającą się ponad większość wartości.

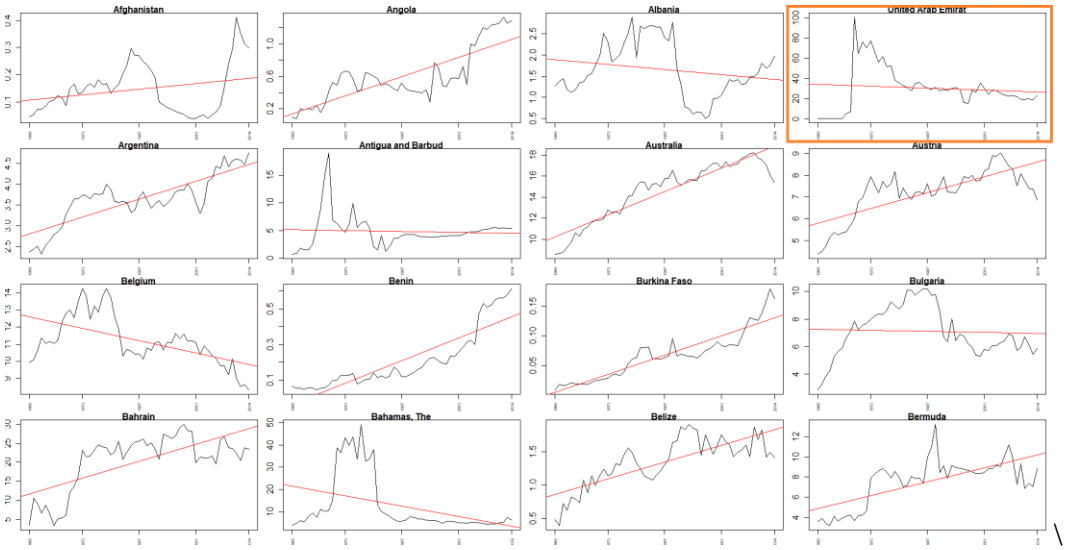


rys. 3.5

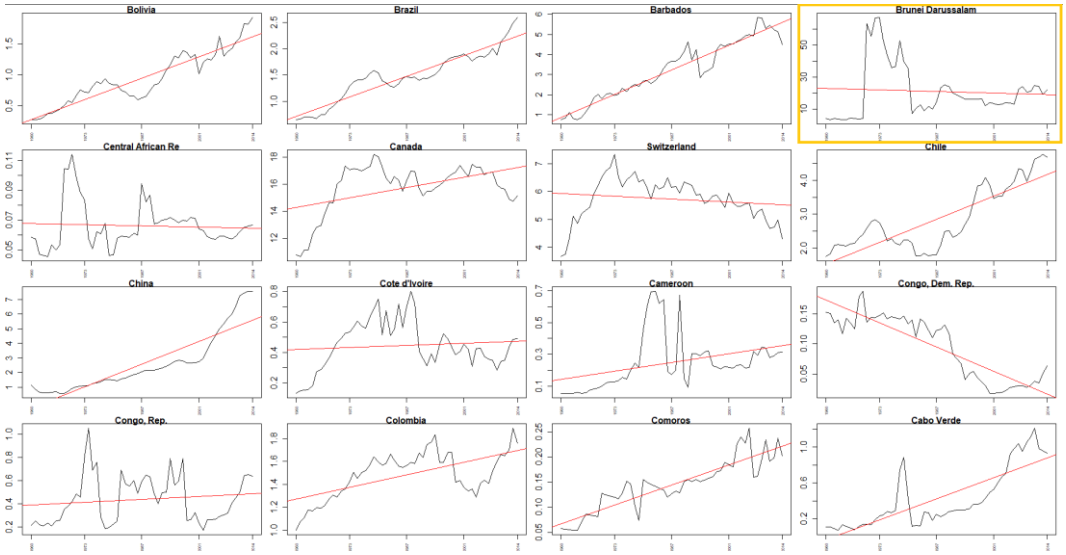
Po wykonaniu osobnych wykresów liniowych i linii trendu (rys. 3.6–15)dla każdego z państw można zauważyć że szukane państwa to: linia fioletowa – Katar, linia pomarańczowa – Zjednoczone Emiraty Arabskie, linia żółta – Brunei, linia niebieska – Luksemburg. W wielu państwach występuje wyraźny trend wzrostu emisji CO2 (np. Wietnam, Benin, Indonezja, Indie).



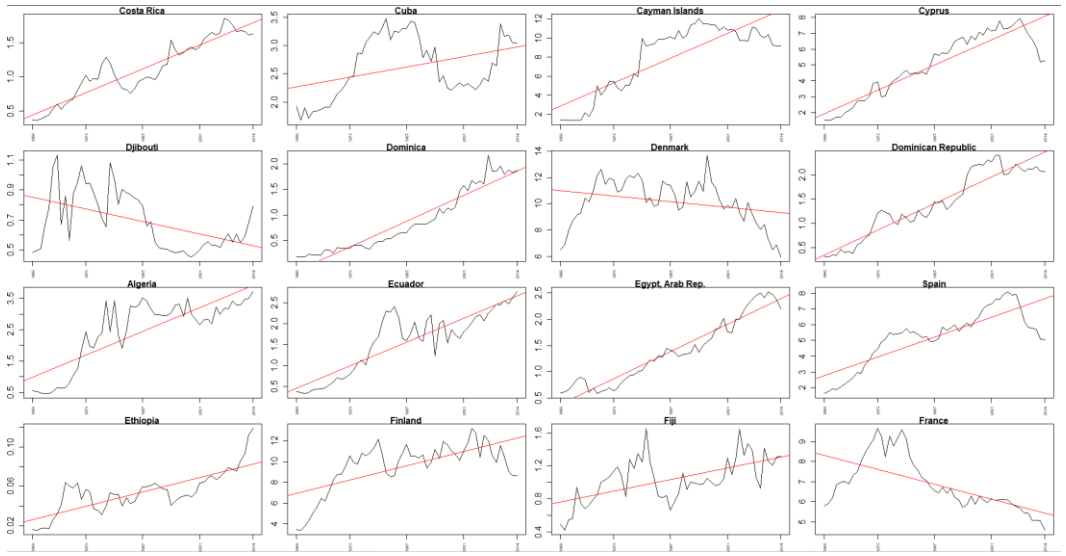
rys. 3.6



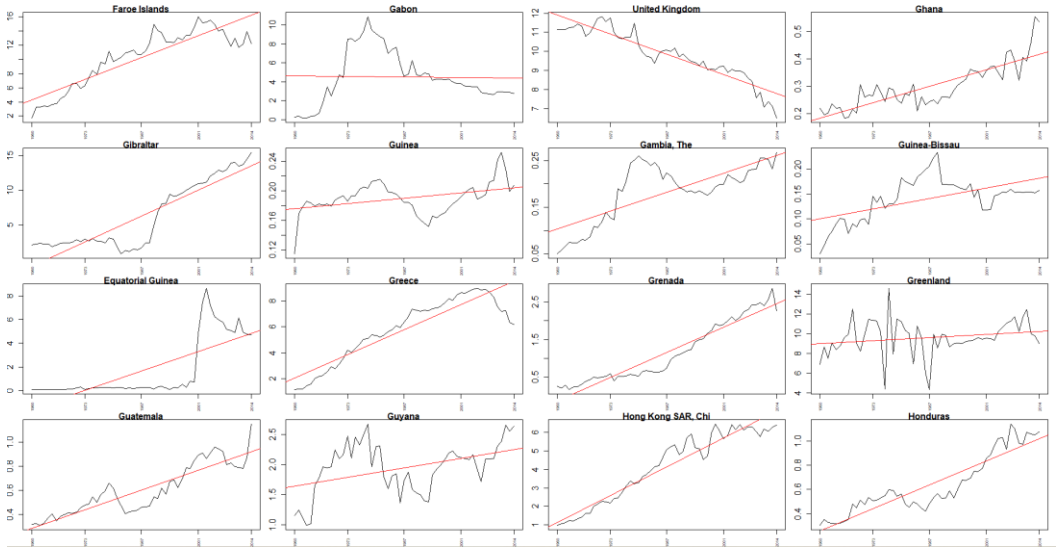
rys. 3.7



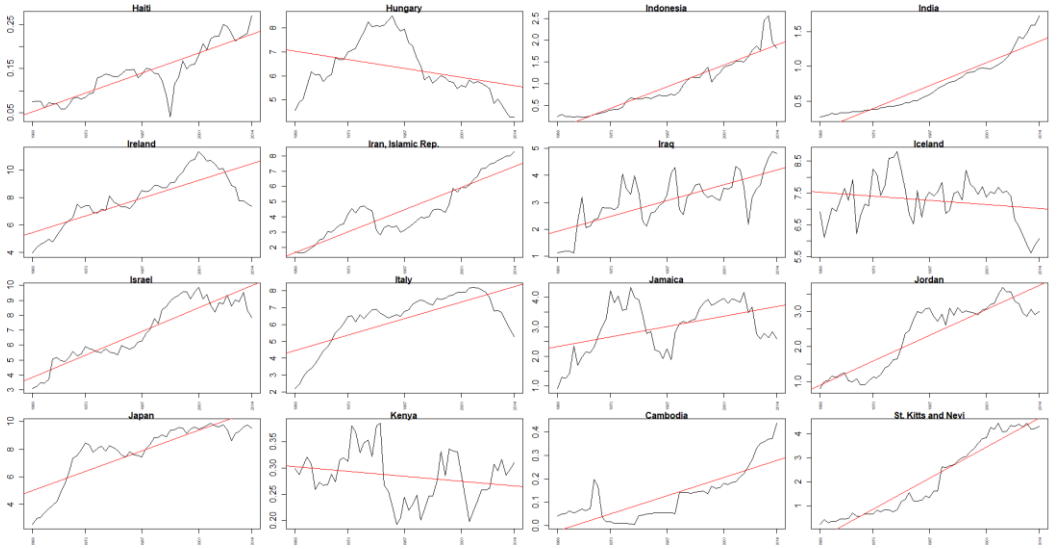
rys. 3.8



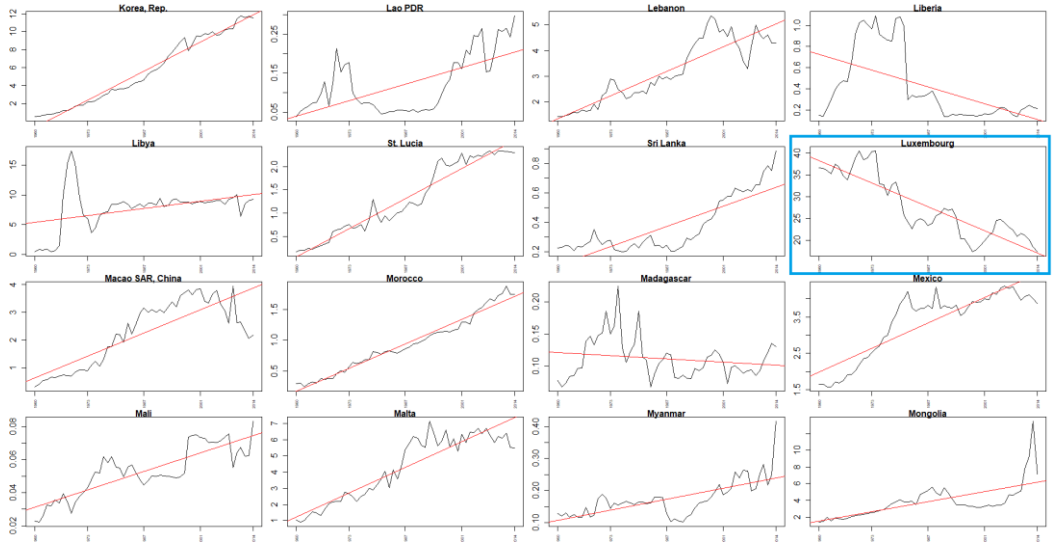
rys. 3.9



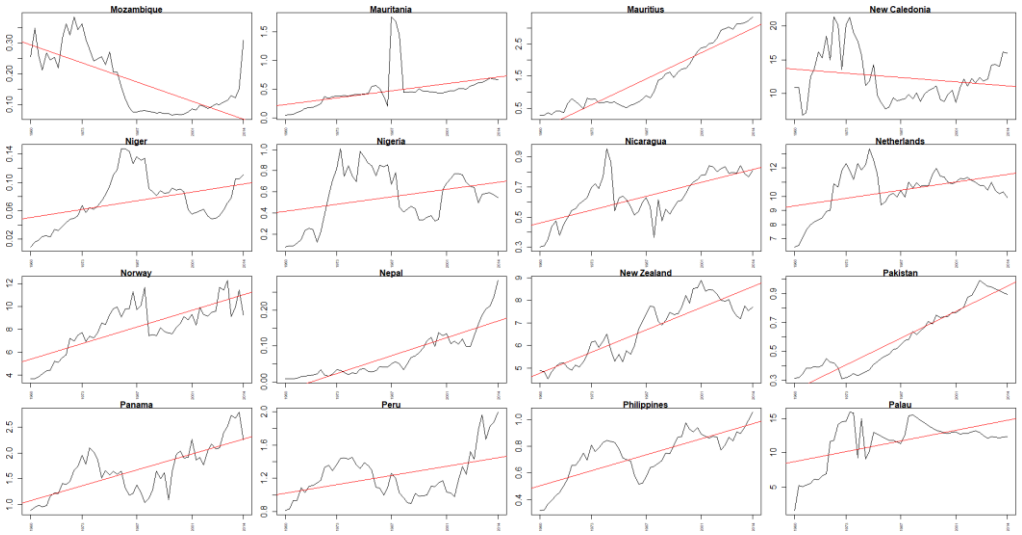
rys. 3.10



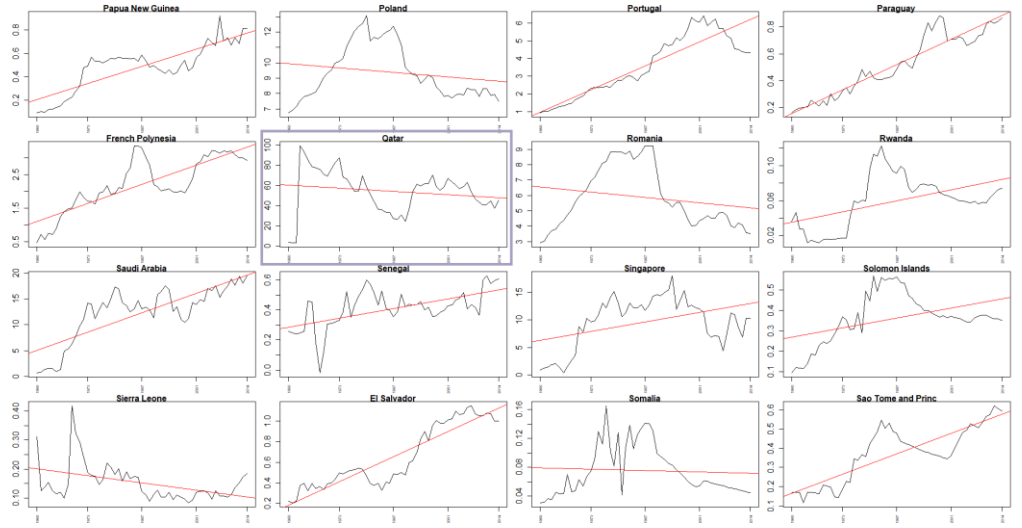
rys. 3.11



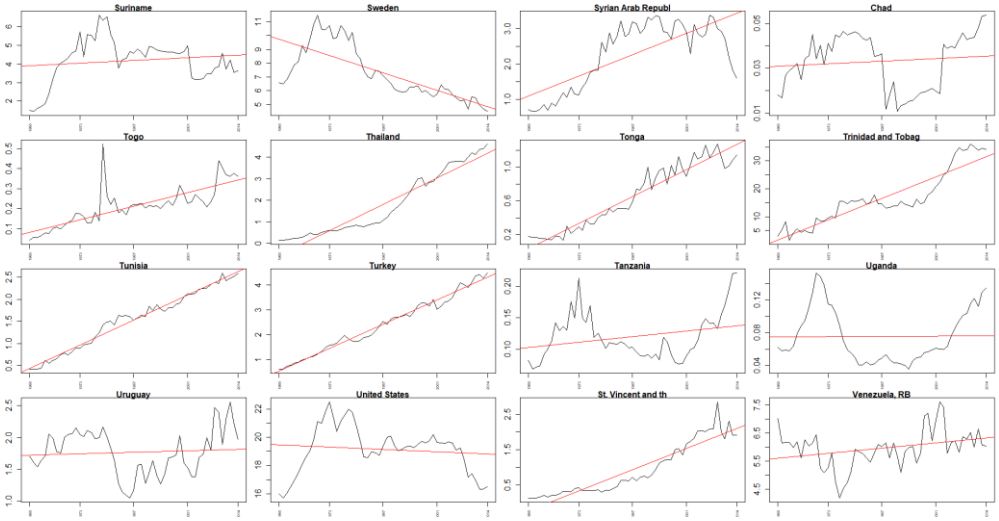
rys. 3.12



rys. 3.13



rys. 3.14

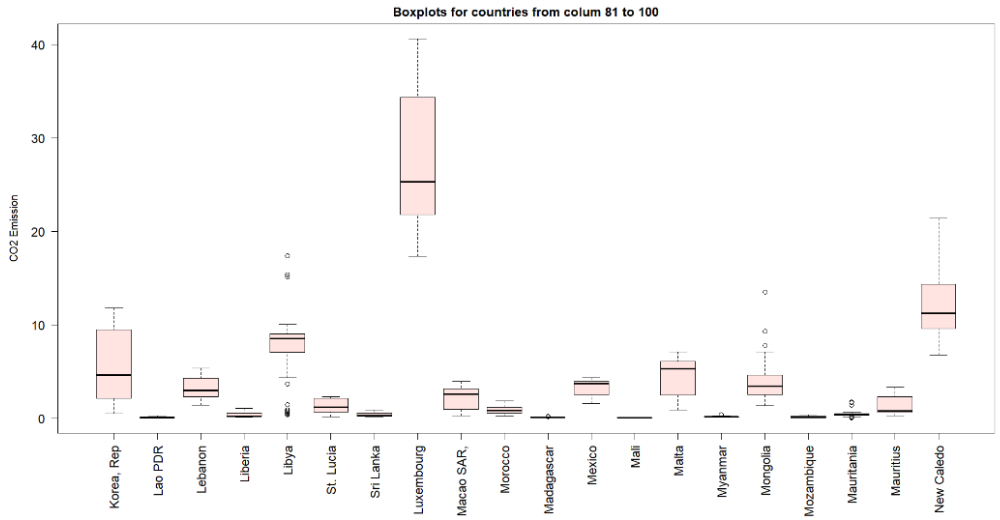


rys. 3.15

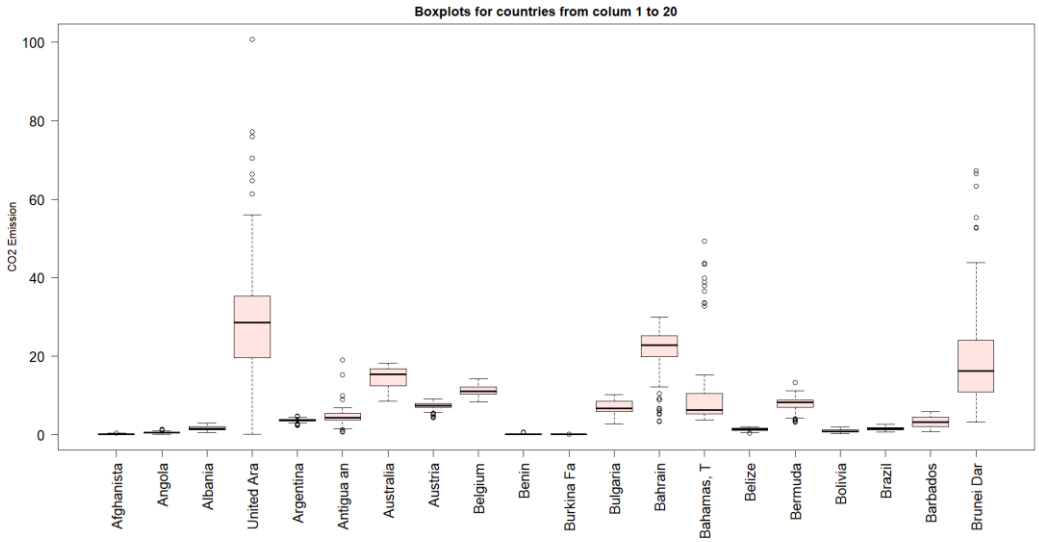
Na wykresach Kataru, Brunei i ZEA widać duży skok wartości w latach 60. Po tym skoku wartości stopniowo maleją do lat 80., nie jest więc to pojedyncza wartość ekstremalna, ale ciąg wyższych wartości trwających przez okres 20 lat. Taka sytuacja może mieć związek ze złożami ropy naftowej, które zaczęły odgrywać najważniejszą rolę w gospodarce tych państw.

Luksemburg natomiast, jest małym państwem o wysokich dochodach, co może wpływać na wysokie wartości wskaźnika CO2, jednak stopniowo malejące na przestrzeni lat.

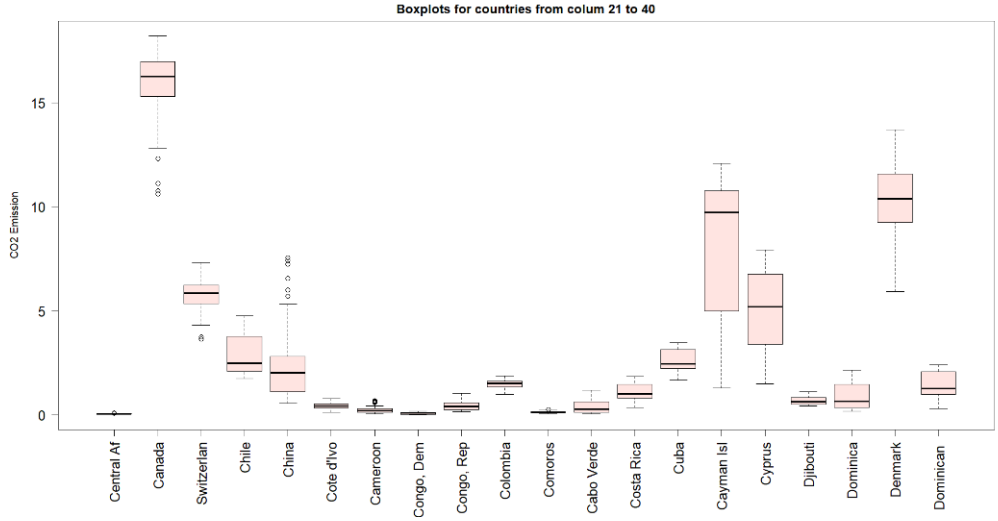
Na rys. od 16 do 24 zaprezentowano boxploty dla każdego z państw, większość z nich nie posiada obserwacji odstających, co oznacza brak poważnych skoków w wysokościach wskaźnika.



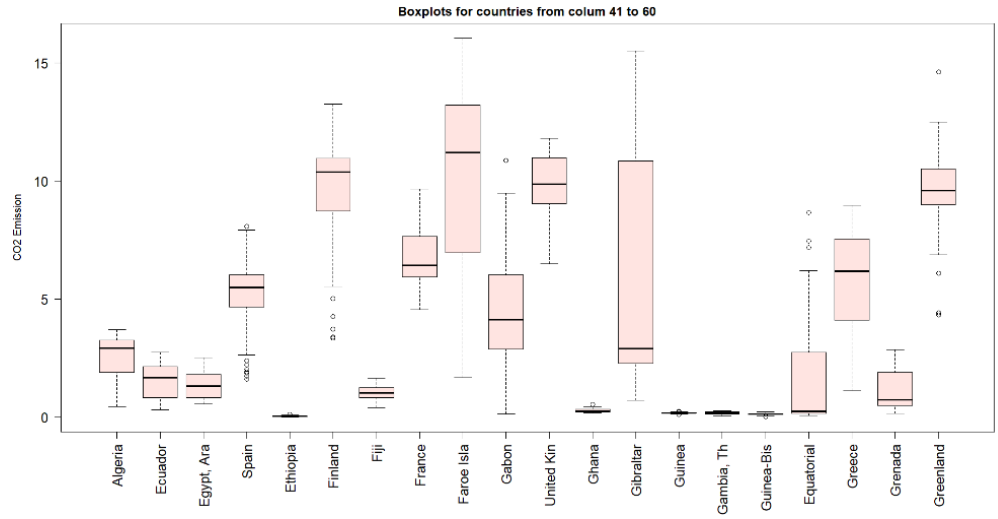
rys. 3.16



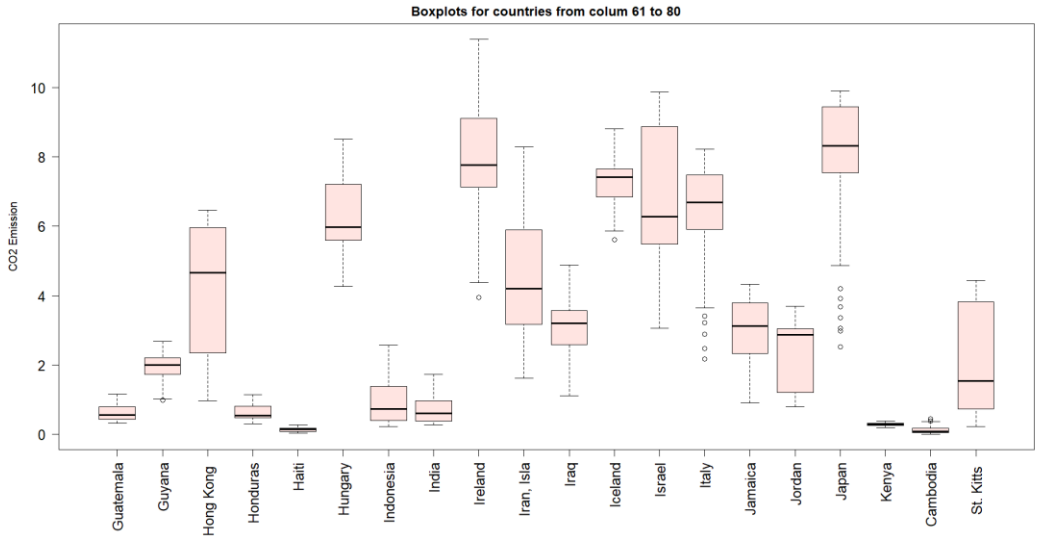
rys. 3.17



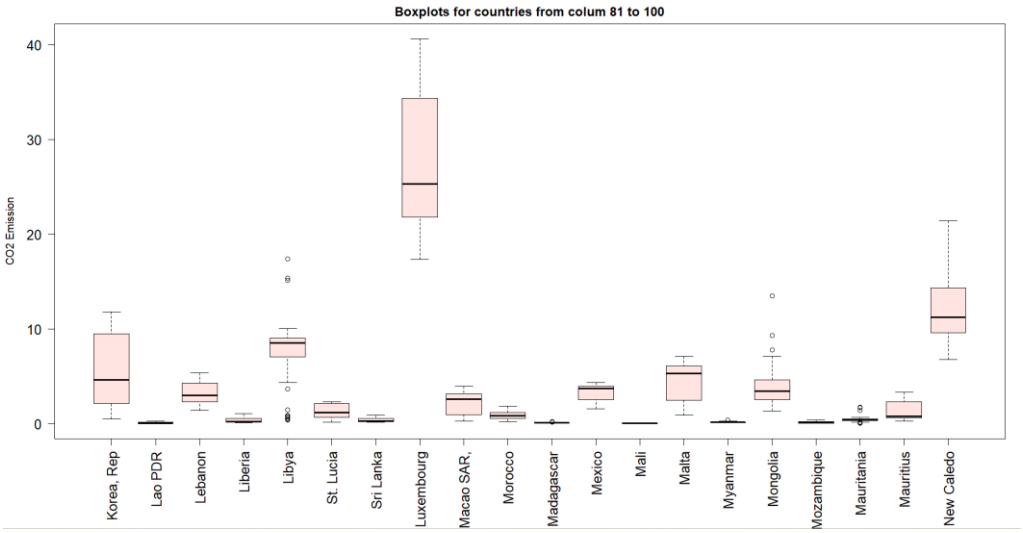
rys. 3.18



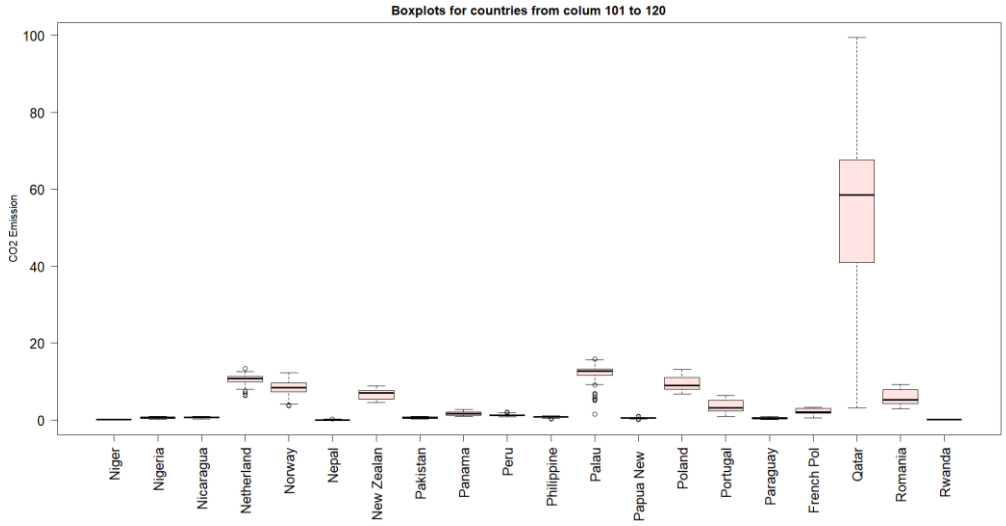
rys. 3.19



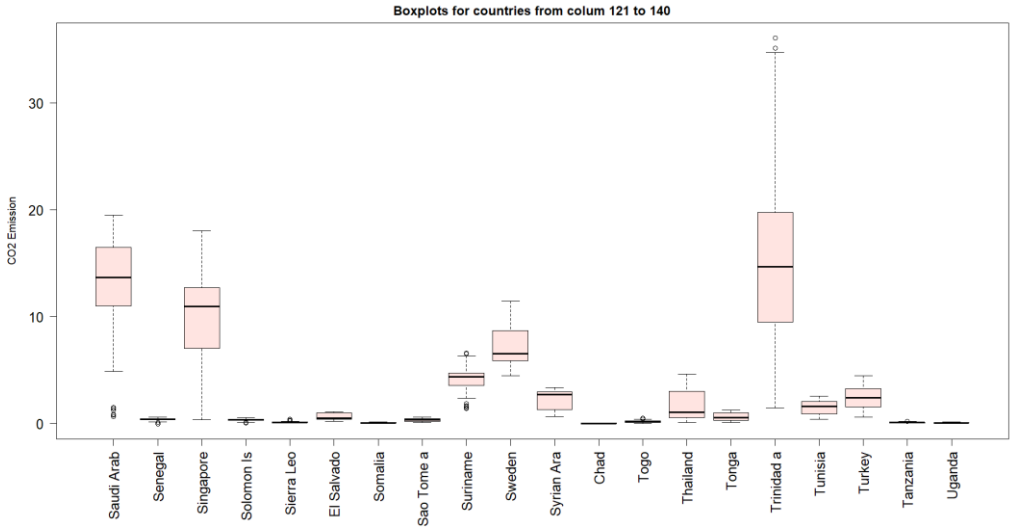
rys. 3.20



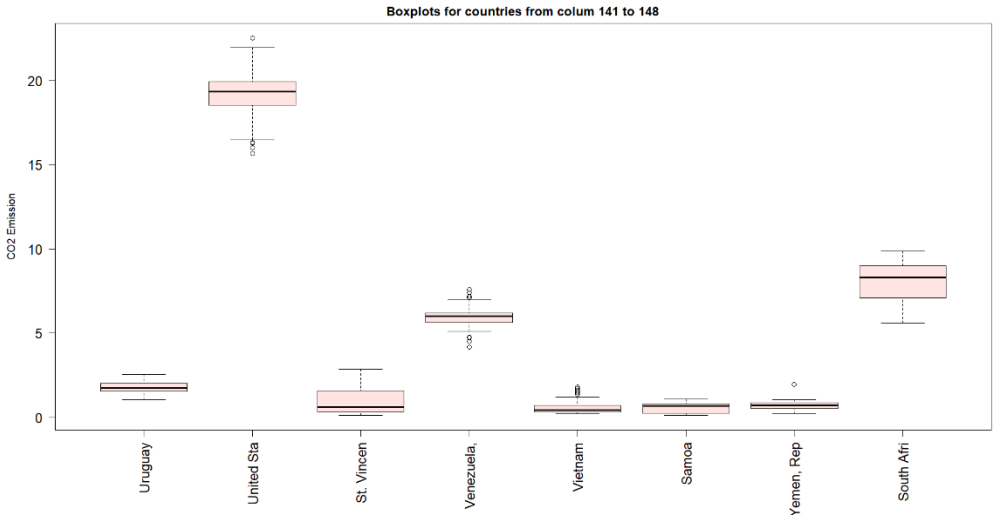
rys. 3.21



rys. 3.22

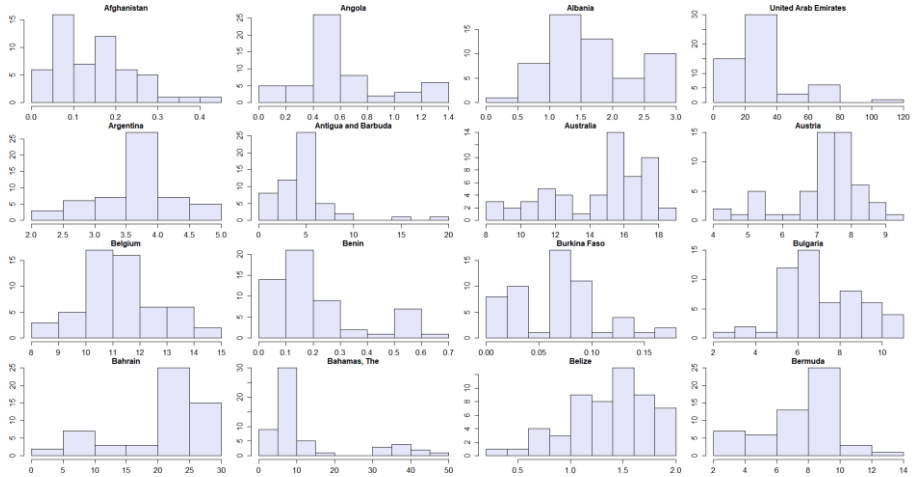


rys. 3.23

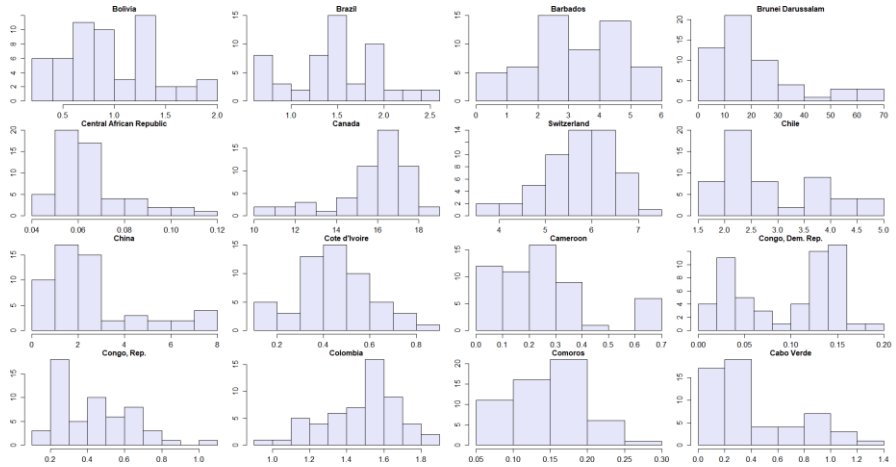


rys. 3.24

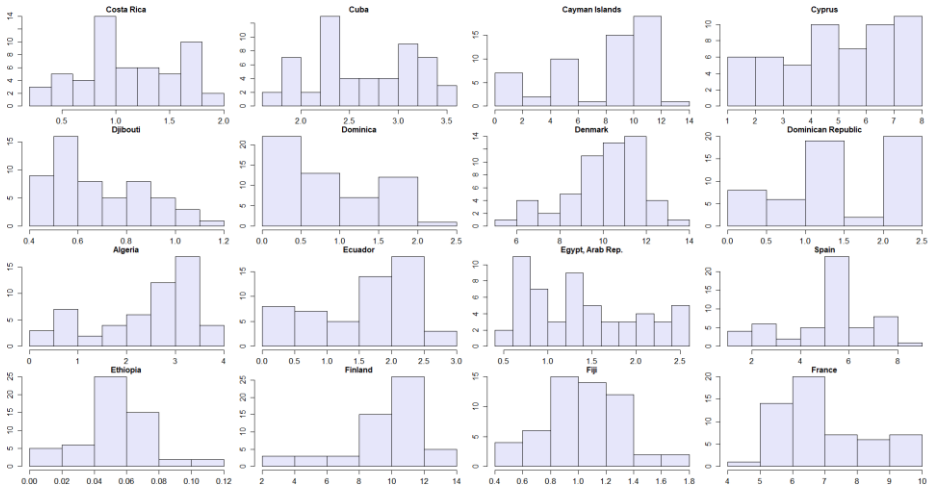
Zostały również przygotowane histogramy (rys. 3.25–34).



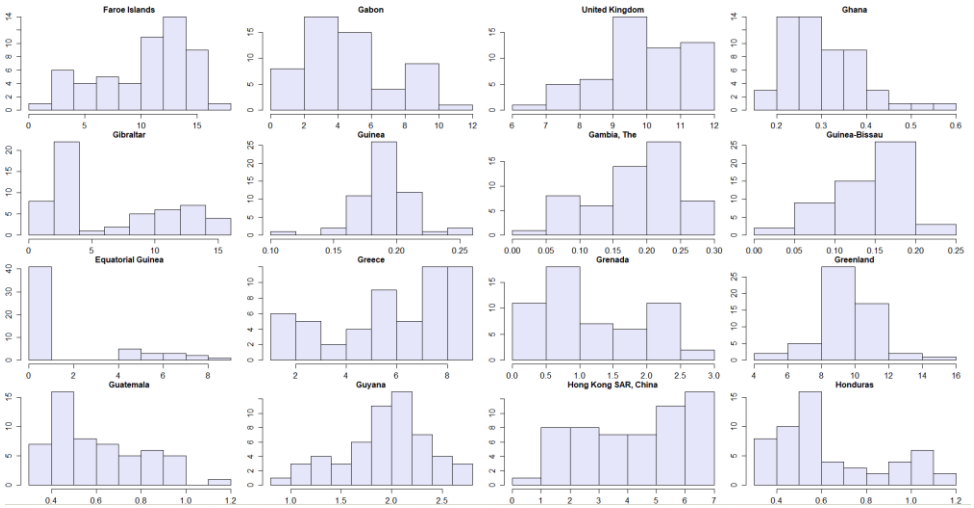
rys. 3.25



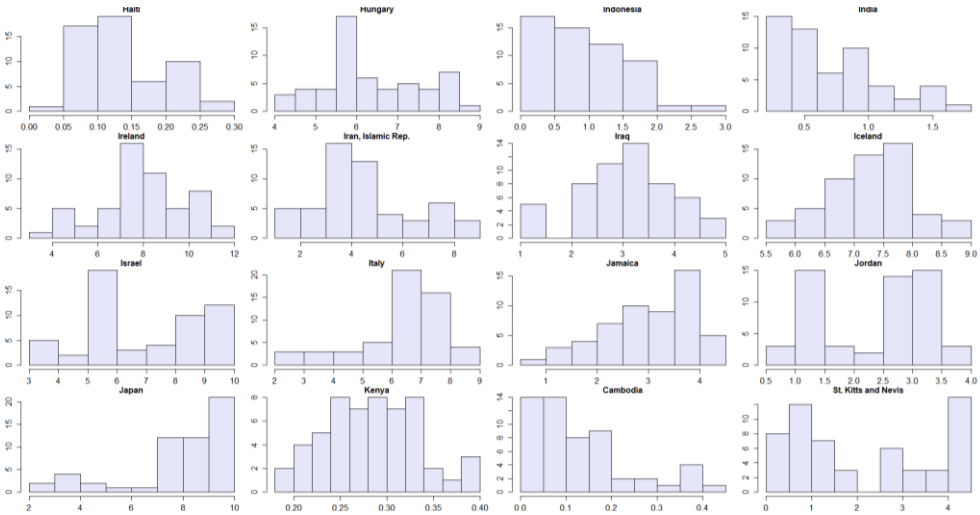
rys. 3.26



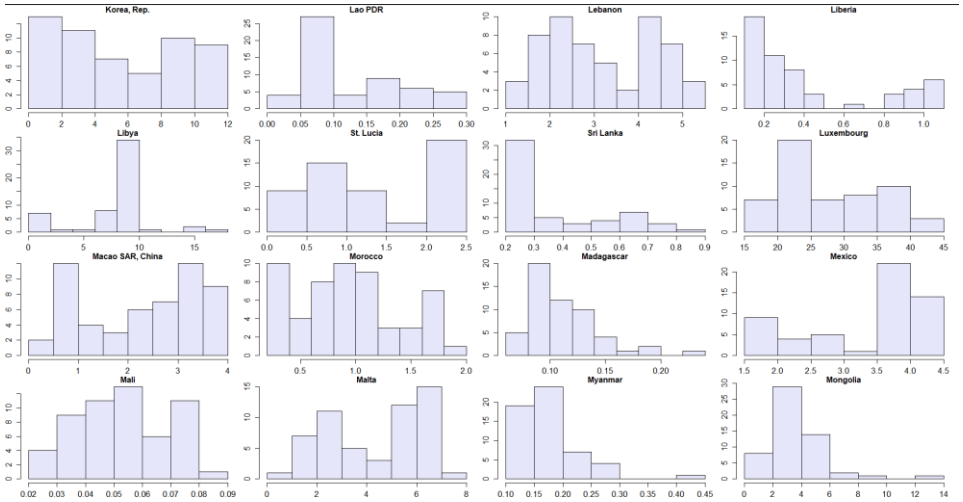
rys. 3.27



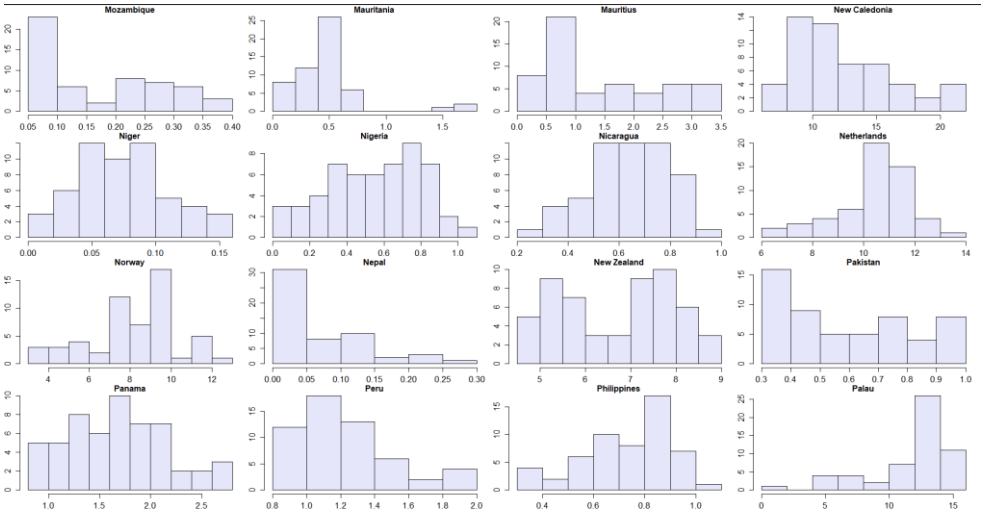
rys. 3.28



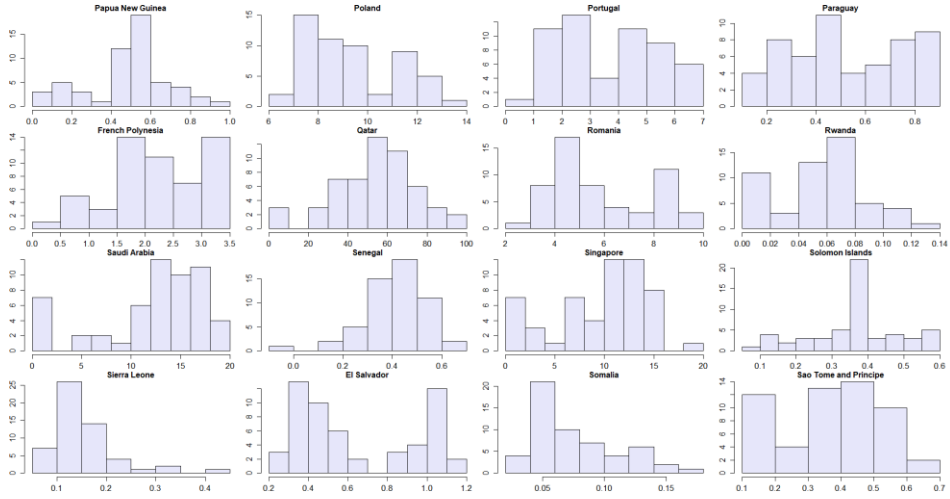
rys. 3.29



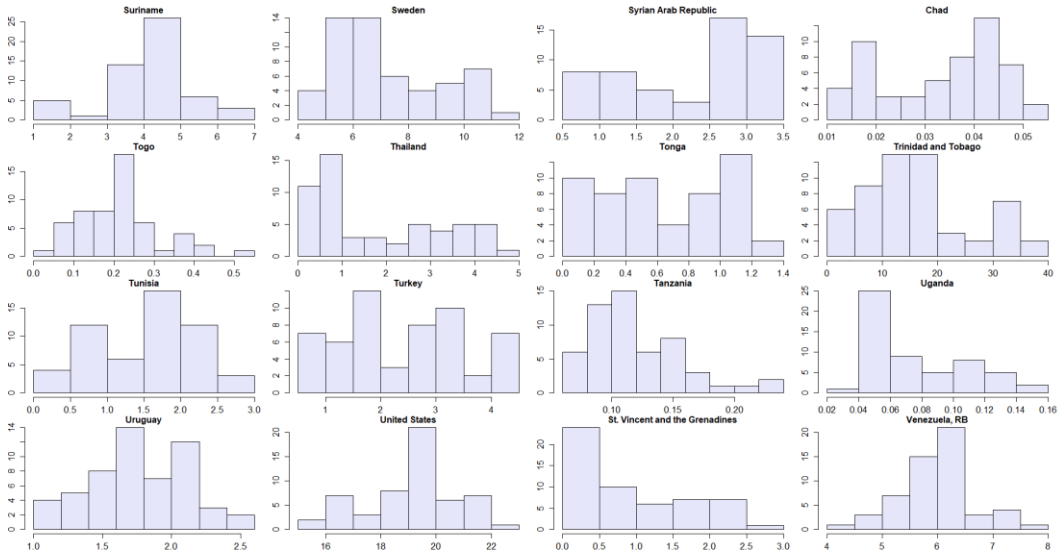
rys. 3.30



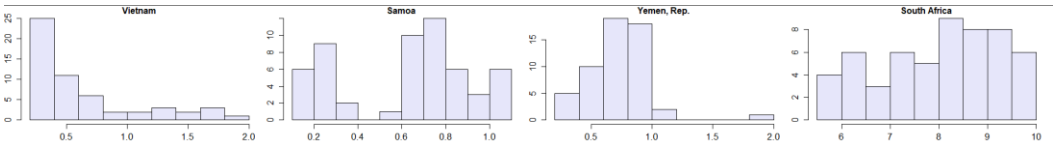
rys. 3.31



rys. 3.32



rys. 3.33



rys. 3.34

4. Model regresyjny Knn

Wylosowany model Knn, w odmianie regresyjnej, został wykorzystany w projekcie ze względu na możliwość przewidywania wartości wskaźnika w ostatnim roku szeregu czasowego. Model ten powinien sobie poradzić z nieliniowymi zmianami występującymi w przypadku większości państw, jest elastyczny i dzięki wykorzystywanej metodzie najbliższych sąsiadów może dostosowywać się do lokalnych zmian.

Aby uzyskać wartości numeryczne w całej ramce danych, dodano dwie nowe kolumny „NumRegion" i „NumIncomeGroup". Kolumna „NumRegion" zawiera klasy od 1 do 7 odpowiadające wartościom opisowym w kolumnie Region, a kolumna „NumIncomeGroup" zawiera klasy od 1 do 4 odpowiadające wartościom opisowym w kolumnie „IncomeGroup".

Po standaryzacji danych przy pomocy funkcji scale() podzielono zbiór na treningowy – 80%, testowy do walidacji – 20% i testowy do predykcji – 4 obserwacje. Następnie wykonano model Knn funkcją knnreg() z pakietu caret (rys. 4.1). Najefektywniejszą wartością parametru k, czyli liczby najbliższych sąsiadów uwzględnianych przy przewidywaniu wartości, była liczba 4.

```
model_knn<- knnreg(train_std_knn, train_y_knn, k = 4)
```

rys. 4.1

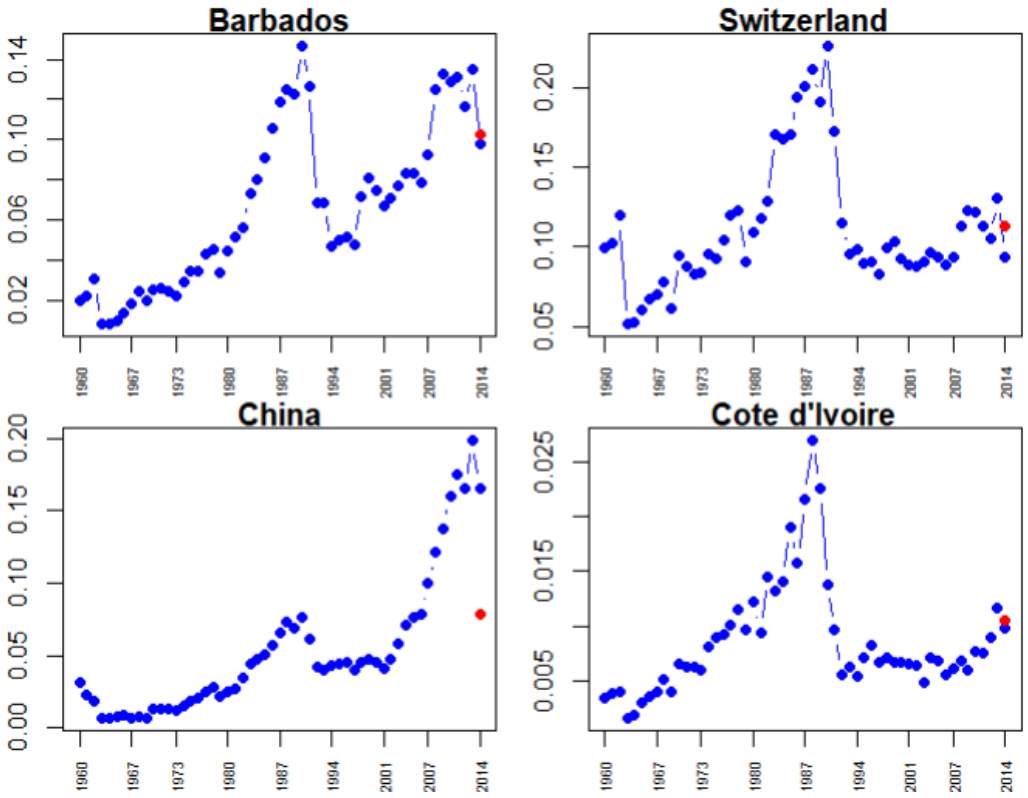
Najpierw wykonano model na większym zbiorze danych testowych do walidacji.

Na rys. 4.2 widać wartości rzeczywiste zestawione z przewidywanymi dla kolejnych wierszy zbioru testowego na danych zestandaryzowanych dla pierwszych czterech wartości.

	Barbados	Szwajcaria	Chiny	Wyb. Kości Słoniowej
real	0.0979	0.0940	0.1652	0.0098
predicted	0.1023	0.1126	0.0782	0.0105

rys. 4.2

Poniżej zaprezentowano wykres porównujący przewidywaną wartość dla roku 2014 (czerwony) do wartości rzeczywistych na przestrzeni lat 1960–2014 (niebieski). Dla 3 na 4 wylosowanych państw, Barbadosu, Szwajcarii i Wybrzeża Kości Słoniowej, można optycznie stwierdzić, że przewidziane wartości były bliskie rzeczywistym. Większa różnica natomiast występowała w przypadku Chin. (rys. 4.3)



rys. 4.3

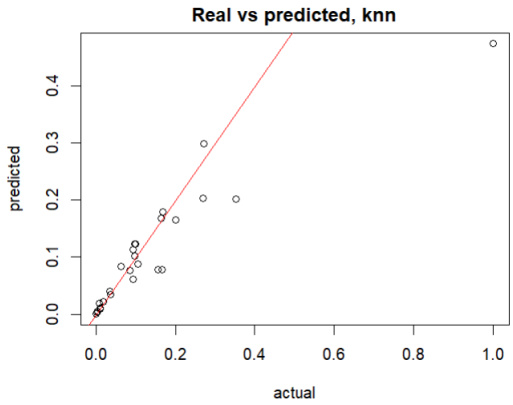
Obliczono również wskaźnik pozwalający ocenić jakość modelu i prognozy. Stosunkowo wysoka wartość współczynnika determinacji R2 wskazuje na dobre dopasowanie modelu do danych. Wartości MSE i RMSE są bardzo niskie. Niepokojąca jest natomiast wysoka wartość MAPE, która sugeruje dużą średnią procentową różnicę między przewidywanymi, a rzeczywistymi wartościami. (rys. 4.4)

test: 20%	R2	MSE	RMSE	MAPE
Knn	0.85	0.01	0.11	25.22%

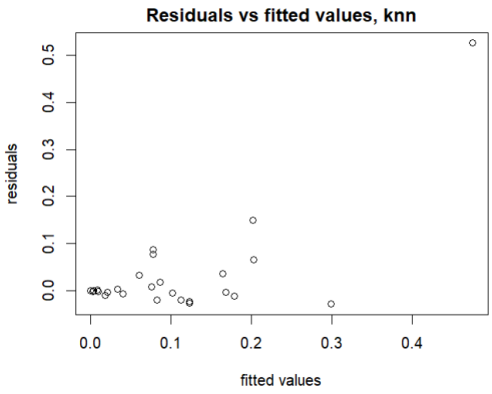
rys. 4.4

Następnie wykonano wykres porównujący rzeczywiste wartości z przewidywanymi (rys. 4.5) oraz wykresy dotyczące reszt: wykres rozrzutu reszt (rys. 4.6), histogram reszt (rys. 4.7) , porównanie rozkładu reszt z rozkładem normalnym (rys. 4.8) , ACF (rys. 4.9) i PACF (rys. 4.10) .

Na wykresie porównujący rzeczywiste wartości z przewidywanymi (rys. 4.5) wartości nie są bardzo oddalone od czerwonej linii, a na wykresie dotyczącym rozrzutu reszt (rys. 4.6) wartości są ułożone mniej więcej losowo. Oba wykresy zaburza jednak mocno odstająca pojedyncza wartość w prawym górnym rogu, która może świadczyć o nagłym skoku wskaźnika, którego model nie przewidział.

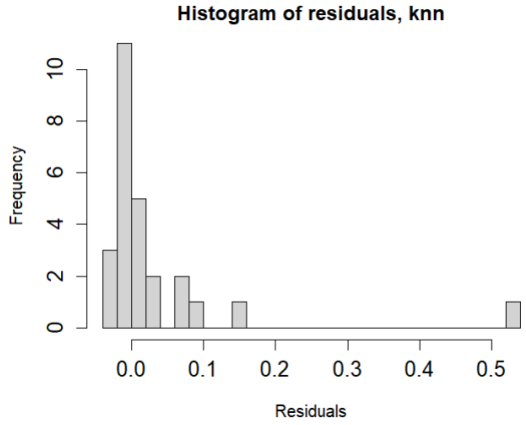


rys. 4.5

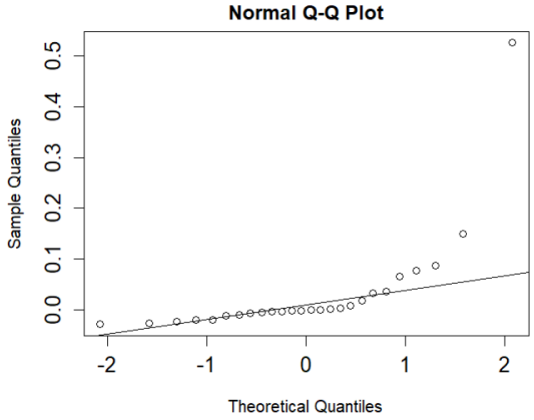


rys. 4.6

Histogram reszt (rys. 4.7) jest prawostronny i nie przypomina pożądanego kształtu rozkładu normalnego. Brak rozkładu normalnego potwierdza wykres porównania rozkładu reszt z rozkładem normalnym (rys. 4.8), gdzie punkty po prawej stronie wykresu są mocno odchylone od linii referencyjnej.

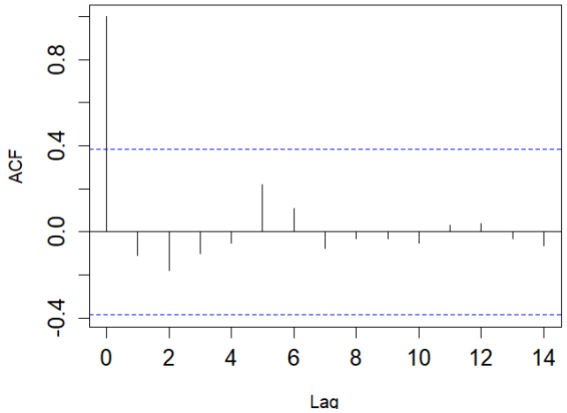


rys. 4.7

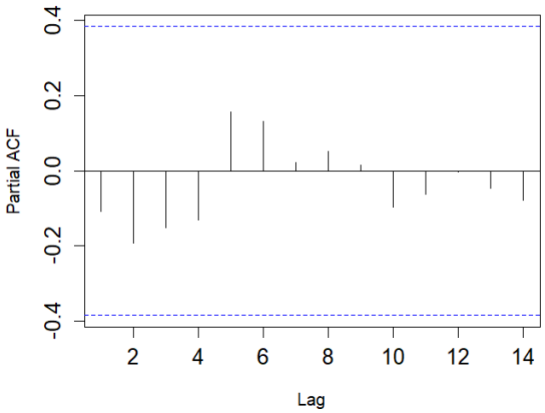


rys. 4.8

Wykresy ACF (rys. 4.9) i PACF (rys. 4.10) potwierdzają brak występowania trendu i sezonowości, pionowe czarne linie nie przebijają poziomych linii niebieskich (poza pierwszą linią na wykresie ACF oznaczającą korelację z samym sobą)



rys. 4.9



rys. 4.10

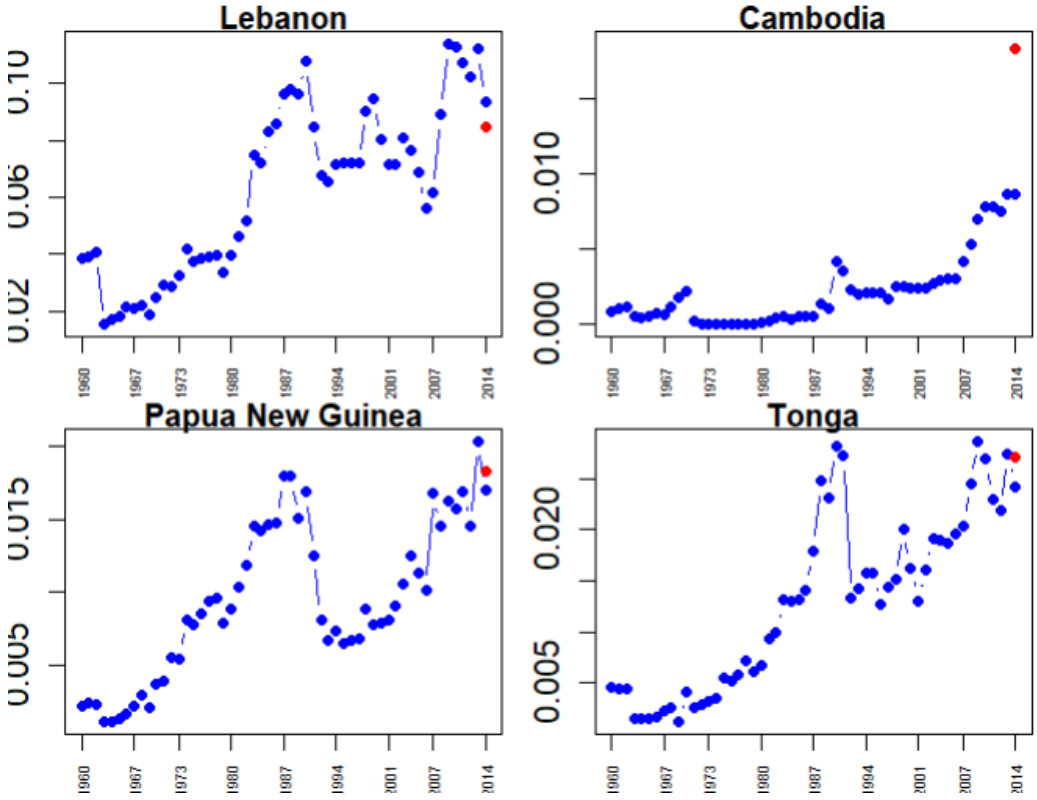
Następnie sprawdzono model na zbiorze testowym do predykcji zawierającym 4 obserwacje.

W poniższej tabelce przedstawionowartości rzeczywiste zestawione z przewidywanymi dla wszystkich państw. (rys. 4.11)

	Liban	Kamboża	Papua Nowa Gwinea	Tonga
real	0.0936	0.0086	0.0169	0.0242
predicted	0.0846	0.0183	0.0183	0.0270

rys. 4.11

Przygotowano wykres porównujący zmianę w czasie danych rzeczywiste z przewidywaną wartością w roku 2014 (rys. 4.12). Bardzo dobrze przewidziane zostały wartości dla Libanu, Papui Nowej Gwinei i Tongy. Natomiast już dla Kambodży wartość przewidywana mocno odbiegała od rzeczywistej.



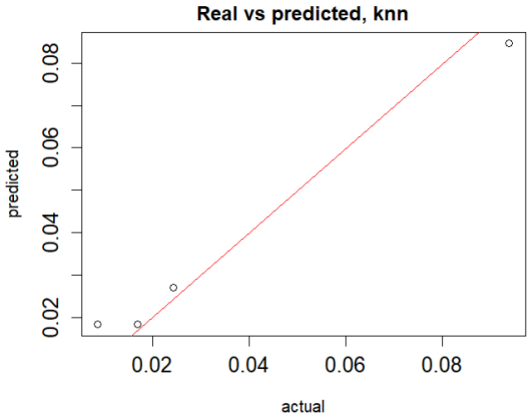
rys. 4.12

W poniższej tabelce (rys. 4.13) zostały przedstawione poprzednio liczone wskaźniki dla nowego podziału. Wartość MSE i za tym RMSE mocno spadły, natomiast współczynnik R2 bardzo wzrósł, co świadczy on o dobrym dopasowaniu modelu do danych. Niepokojący jest natomiast wzrost MAPE do aż 35%.

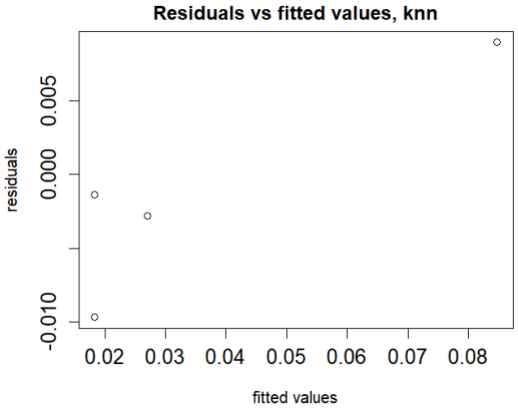
test: 4	R2	MSE	RMSE	MAPE
Knn	0.99	4.61E-05	0.0067	35.25%

rys. 4.13

Na wykresie porównującym rzeczywiste wartości z przewidywanymi (rys. 4.14) widać, że punkty układają się stosunkowo blisko przekątnej. Punkty na wykresie rozrzutu reszt (rys. 4.15) są rozrzucone losowo i nie układają się w żaden konkretny kształt.

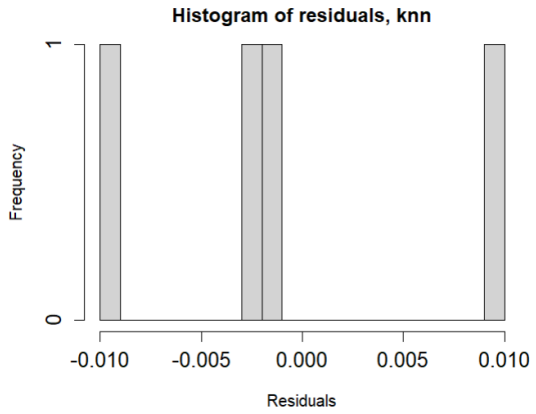


rys. 4.14

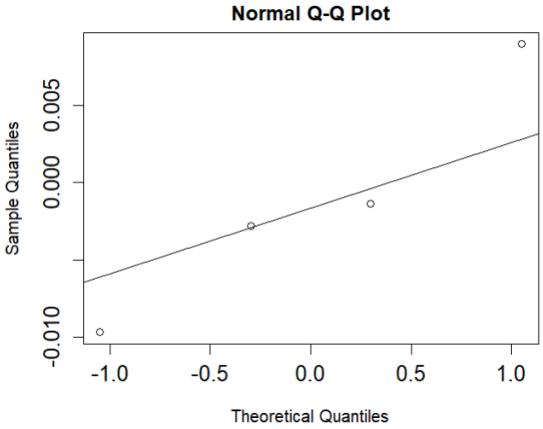


rys. 4.15

Histogram reszt (rys. 4.16) nie przypomina rozkładu normalnego jednak wpływ na to może mieć mała ilości obserwacji. Na wykresie porównania rozkładu reszt z rozkładem normalnym (rys. 4.17) odchylenia na krańcowych wartościach oznaczają większą ilość wartości ekstremalnych niż w przypadku występowania prawdziwego rozkładu normalnego.

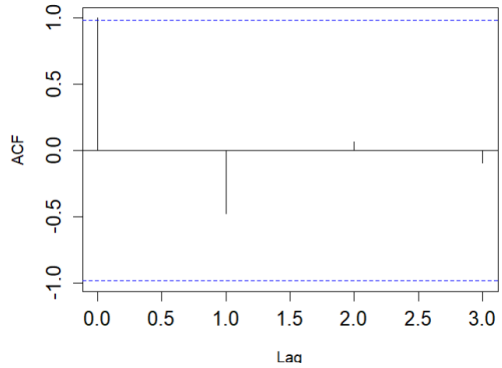


rys. 4.16

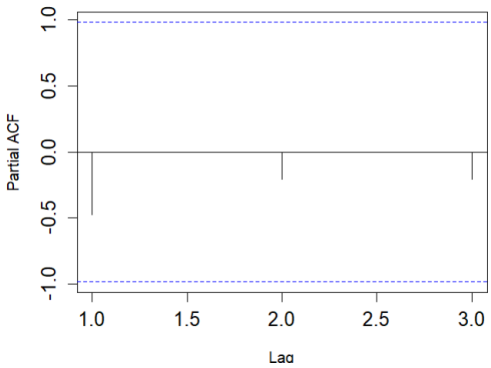


rys. 4.17

Ponownie wykresy ACF (rys. 4.18) i PACF (rys. 4.19) potwierdzają brak występowania trendu i sezonowości.



rys. 4.18



rys. 4.19

Po analizie danych stwierdzono, że odstający punkt reprezentuje Katar, który charakteryzuje się dużymi zmianami wskaźnika. Pomimo braku poprawnej predykcji w tym punkcie, model bardzo dobrze poradził sobie z predykcją dla państw z mniejszymi zmianami wartości, co potwierdzają wysokie wartości R2, niskie MSE i wizualnie wykresy z nałożonymi wartościami rzeczywistymi i przewidywanymi (rys, 4.3, rys. 4.12).

5. Model regresyjny sieci neuronowych MLP

Do wykonania drugiego modelu wybrano sieci neuronowe MLP z uwagi na możliwość wykorzystania warstw ukrytych, które mogą przetwarzać złożone dane. Model jest również elastyczny i może dostosowywać się do zmieniających się warunków.

Ponownie podzielono zestandaryzowane dane na trzy zbiory, treningowy, walidacyjny i do predykcji. Do wykonania modelu użyto funkcji neuralnet z dwoma ukrytymi warstwami: pierwszą z czterema neuronami i drugą z trzema. (rys. 5.1)

```
model_mlp <- neuralnet(year2014 ~ . ,
                       data = data_train_mlp, hidden = c(4, 3),
                       linear.output = TRUE)
```

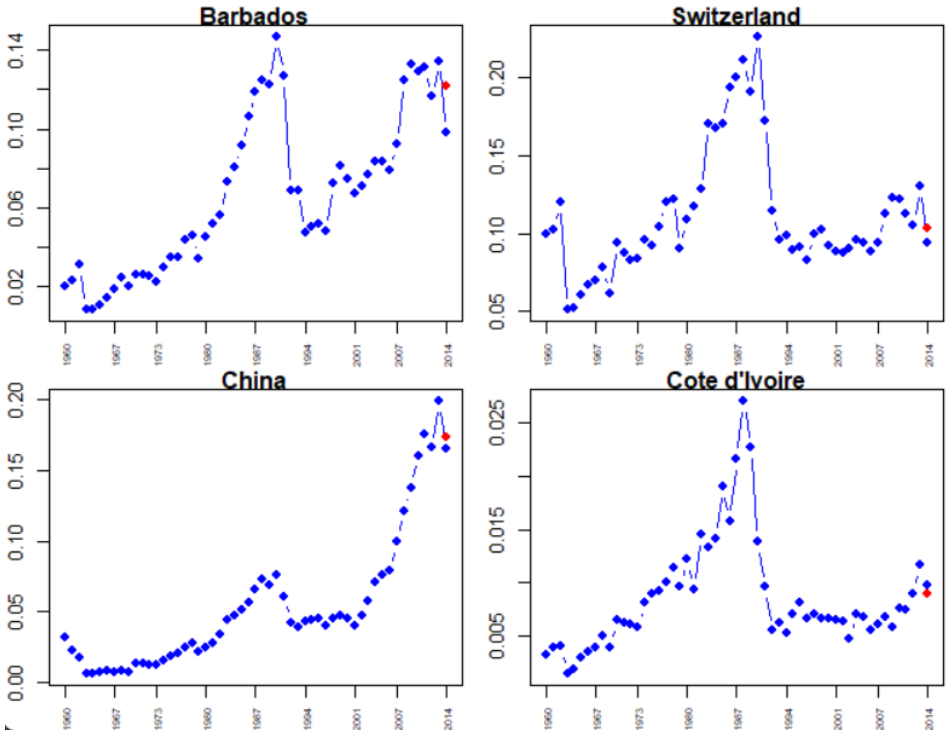
rys. 5.1

Na rys. 5.2 przedstawiono wartości przewidywane i rzeczywiste dla pierwszych czterech państw ze zbioru testowego do walidacji.

	Barbados	Szwajcaria	Chiny	Wyb. Kości Słoniowej
real	0.0979	0.0940	0.1652	0.0098
predicted	0.1217	0.1038	0.1734	0.0089

rys. 5.2

Następnie przygotowano wykres, dzięki któremu można porównać wizualnie jak zostały przewidziane wartości dla roku 2014. W każdym przypadku predykowane wartości są bliskie rzeczywistym (rys. 5.1). Porównując rys. 5.3 do zestawu wykresów na rys. 4.3 wykonanych przy pomocy algorytmu Knn można stwierdzić, że wartości zostały lepiej przyporządkowane przez algorytm MLP.



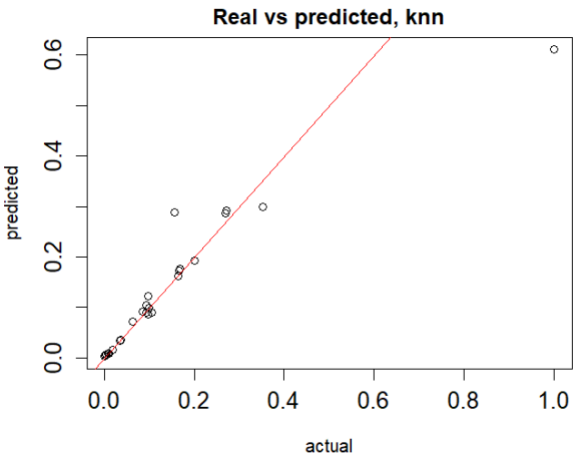
rys. 5.3

Na podstawie tabelki na rys. 5.4 można stwierdzić, że model jest stosunkowo dobrze dopasowany do danych (R2 równe 0.88). MSE i RMSE są zadowalająco niskie, lecz współczynnik MAPE bardzo wysoki.

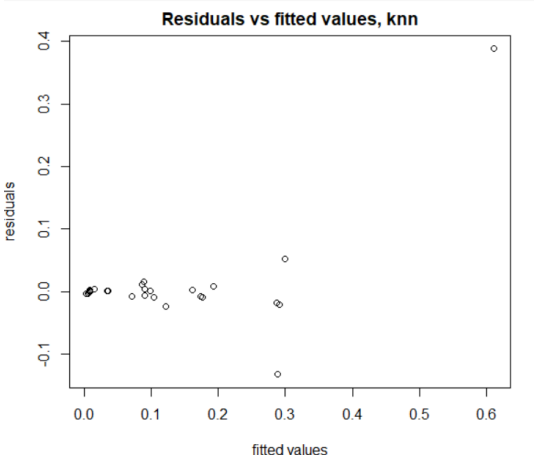
test: 20%	R2	MSE	RMSE	MAPE
Knn	0.88	0.006	0.081	33.93%

rys. 5.4

Ponownie wartości na wykresie porównującym przewidywane i rzeczywiste (rys. 5.5) wartości leżą blisko czerwonej linii, z wyjątkiem jednej odstającej wartości. Na wykresie rozrzutu reszt (rys. 5.6) powtarza się sytuacja z tą samą jedną obserwacją, która mocno odbiega od innych. Reszta wartości jest rozrzucona losowo wokół poziomu zera na osi reszt.

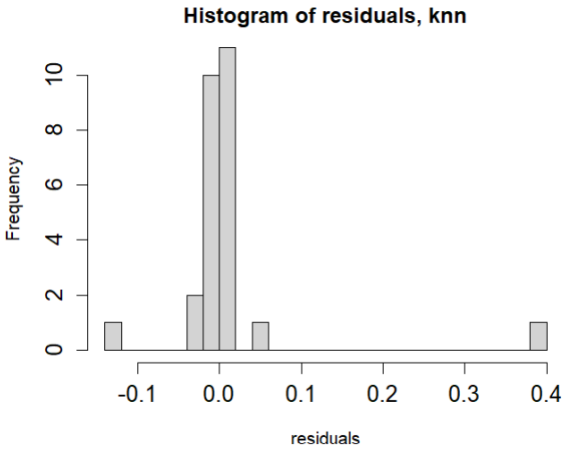


rys. 5.5

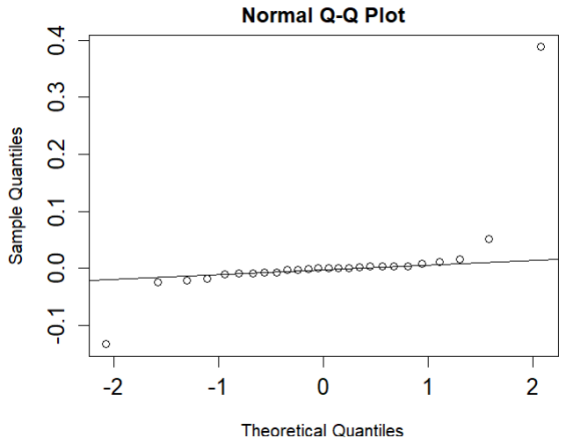


rys. 5.6

Histogram reszt (rys. 5.7) zaburza jedna odstająca wartość, poza tym może on w przybliżeniu przypominać rozkład normalny. Potwierdza to wykres porównania rozkładu reszt z rozkładem normalnym, gdzie poza krańcową wartością, większość obserwacji leży na linii referencyjnej (rys. 5.8).

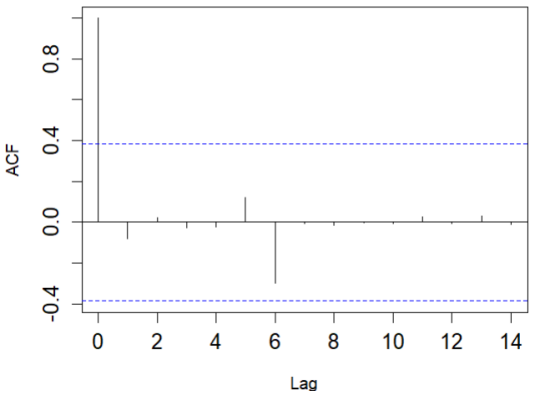


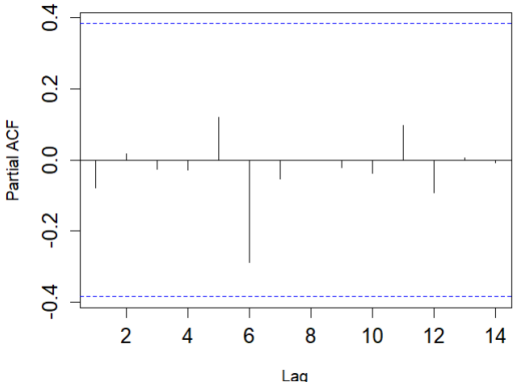
rys. 5.7



rys. 5.8

Wygląd wykresów ACF (rys. 5.9) i PACF (rys. 5.10) świadczy o braku sezonowości i trendu.





rys. 5.9

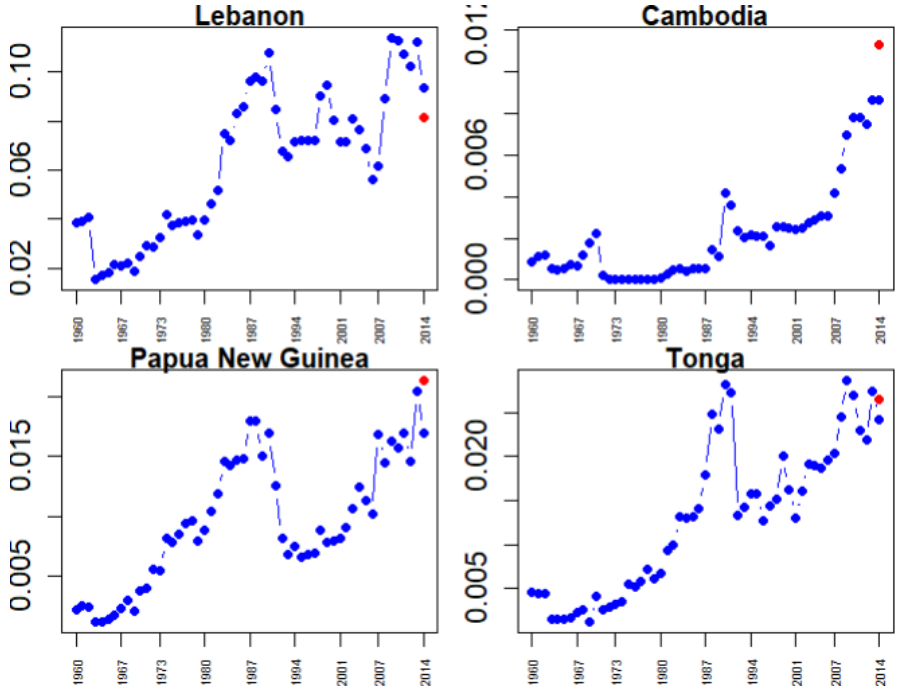
rys. 5.10

Następnie sprawdzono model na zbiorze przeznaczonym do predykcji zawierającym 4 obserwacje. Na rys. 3.11 porównano wartość predykowaną przez model do rzeczywistych wartości dla każdego państwa.

	Liban	Kamboża	Papua Nowa Gwinea	Tonga
real	0.0936	0.0086	0.0169	0.0242
predicted	0.0813	0.0112	0.0213	0.0265

rys. 5.11

Ponownie przygotowano wykres na którym wizualnie można stwierdzić jak dobrze wartości zostały przewidziane przez model. W przypadku Libanu i Tongy wartości zostały przewidziane stosunkowo poprawnie, dużo gorzej jednak oszacowano wartości dla Kambodży i Papui Nowej Gwinei. (rys. 5.12)



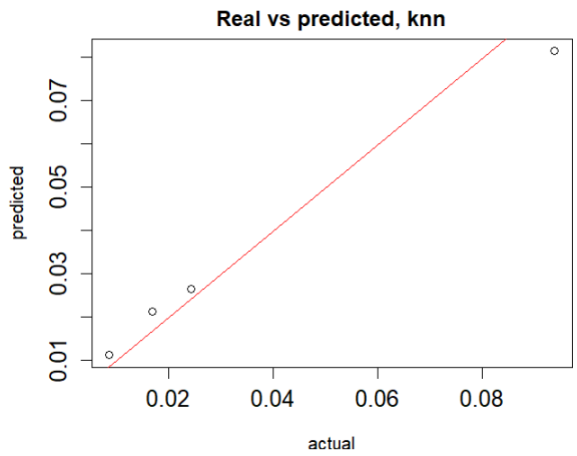
rys. 5.12

Wskaźnik R2 ponownie jest bardzo wysoki, a MSE utrzymuje się na niskim poziomie. Warto zwrócić uwagę na MAPE, które spadło poniżej 19%. (rys. 5.13)

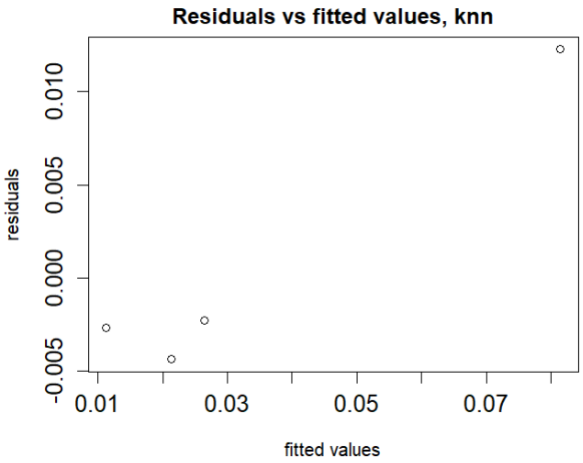
test: 4	R2	MSE	RMSE	MAPE
Knn	0.99	4.56E-05	0.0067	19.00%

rys. 5.13

Wszystkie wartości na wykresie porównującym rzeczywiste i przewidywane (rys. 5.14) wartości leżą blisko linii, a na wykresie rozrzutu reszt obserwacje są rozłożone losowo (rys. 5.15).

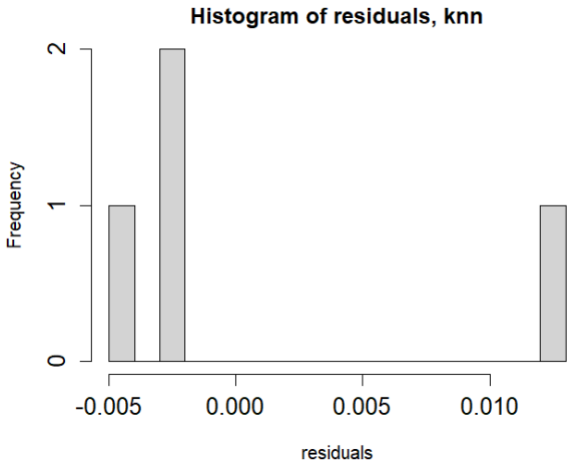


rys. 5.14

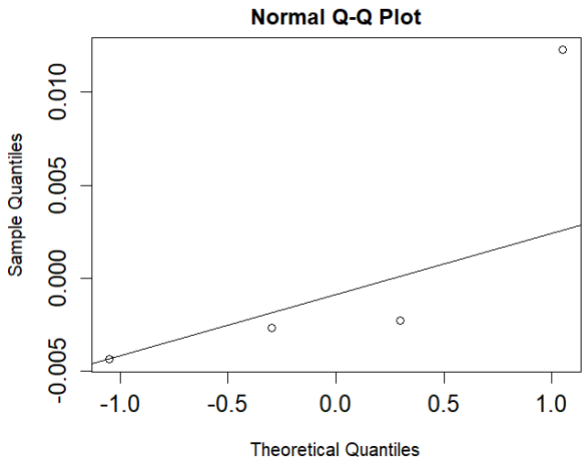


rys. 5.15

Histogram reszt (rys. 5.16) nie ma pożądanego kształtu rozkładu normalnego przez bardzo duże odsunięcie w prawo jednej wartości, obserwację tą potwierdza wykres porównania rozkładu reszt z rozkładem normalnym (rys. 5.17), gdzie wartości leżą w dość dużym oddaleniu od linii referencyjnej.

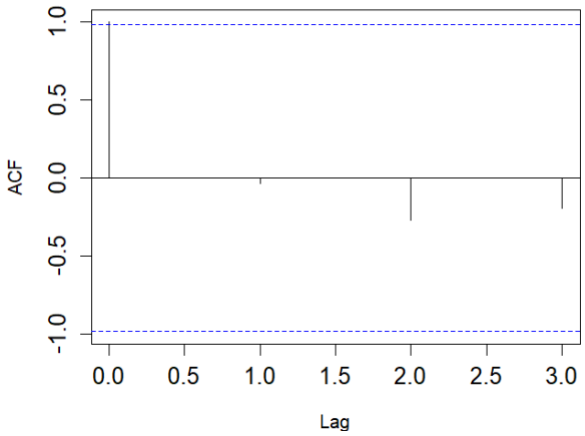


rys. 5.16

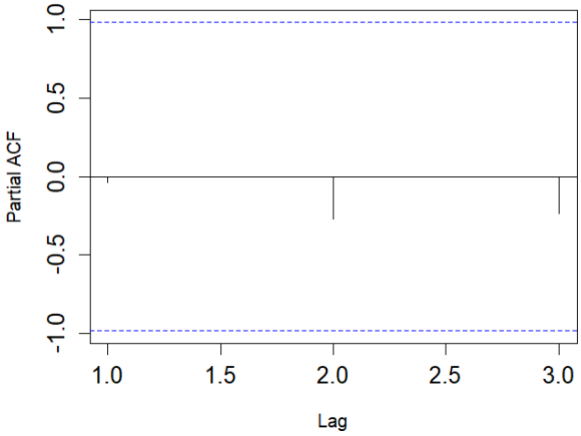


rys. 5.17

Tak jak w każdym poprzednim przypadku wykresy ACF (rys. 5.18) i PACF (rys. 5.19) informują o braku trendu i sezonowości.



rys. 5.18



rys. 5.19

6. Porównanie modeli

W celu porównania modeli wykonanych przy wykorzystaniu algorytmów Knn regresyjnego i sieci neuronowych MLP zestawiono dane z przetestowania modeli na zbiorze do predykcji zawierającym cztery obserwacje.

model	R2	MSE	RMSE	MAPE
Knn	0.99	4.61E-05	0.0067	35.25%
MLP	0.99	4.56E-05	0.0067	19.00%

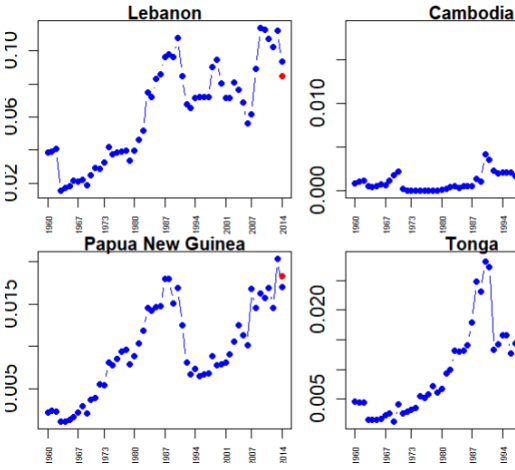
rys. 6.1

Wskaźnik R2 był identyczny dla obu modeli i osiągnął niezwykle wysoką wartość 0.99, co oznacza bardzo dobre dopasowanie modelu do danych. Również nie widać znaczącej różnicy między wskaźnikami MSE (i RMSE), w obu przypadkach były one niskie. Największą różnicę między modelami można zaobserwować porównując wartość MAPE, zdecydowanie niższe było ono dla modelu wykonanego przez MLP, aż o 16 punktów procentowych.

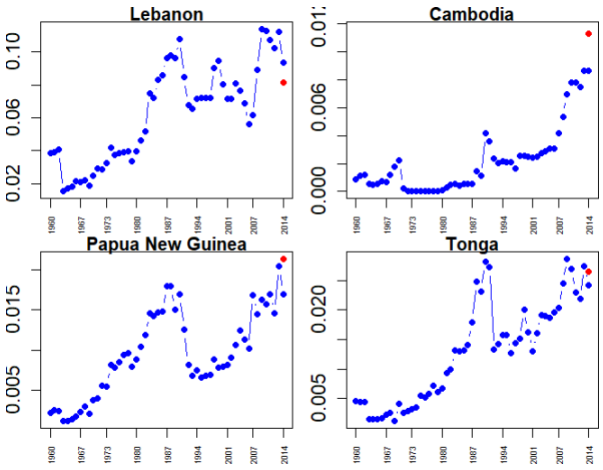
	Liban	Kamboża	Papua Nowa Gwinea	Tonga
real	0.0936	0.0086	0.0169	0.0242
Knn	0.0846	0.0183	0.0183	0.0270
MLP	0.0813	0.0112	0.0213	0.0265
różnica bzw. Knn	0.0090	0.0097	0.0014	0.0028
różnica bzw. MLP	0.0123	0.0026	0.0044	0.0023
zgodnie z kierunkiem Knn	tak	tak	tak	tak
zgodnie z kierunkiem MLP	tak	tak	nie	tak

rys. 6.2

Na rys. 6.2 porównano wartości przewidywane przez oba modele do wartości rzeczywistych, następnie obliczono wartości bezwzględne ich różnic. Model Knn okazał się być lepszy w pierwszym i trzecim przypadku, a model MLP w drugim i czwartym przypadku – nie można na tej podstawie wyciągnąć wniosków o większej skuteczności któregoś z modeli. Jednak kierunek zmian (tzn. wzrost lub spadek wartości w 2014 względem 2013) przewidział poprawnie w każdym przypadku Knn, a MLP pomylił się raz. Zmiana wartości rzeczywistych na przestrzeni lat oraz wartość przewidywana w 2014 zostały przedstawione wizualnie na rys. 6.3 i rys. 6.4.



rys. 6.3



rys. 6.4

7. Wniosek

Oba modele wykonane przez algorytmy Knn regresyjny i sieci neuronowych MLP osiągnęły podobną skuteczność w przewidywaniu wartości wskaźnika emisji CO2 w zależności od państwa w 2014 roku.

Jako nieco lepszy można wskazać algorytm MLP, przemawiają za tym: niższa wartość wskaźnika MAPE i lepszy układ histogramu reszt oraz wykresu porównania rozkładu reszt z rozkładem normalnym.