

Sprawozdanie

Diagnostyka systemów

Prognoszowana długość życia pacjentów na podstawie wskaźników
medycznych

Tomasz Lewandowski

172412

1. Cel projektu

Celem projektu była implementacja modelu pozwalającego na prognozę długości życia pacjentów na podstawie wskaźników medycznych udostępnionych przez WHO (*World Health Organization*).

2. Rozwiązanie

W celu prognozowania długości życia pacjentów zaimplementowany został model wielokrotnej regresji liniowej (MLR). Jest to technika stosowana do przewidywania wyniku zmiennej na podstawie wartości dwóch lub więcej zmiennych. W algorytmie wykorzystywana jest wielokrotna regresja liniowa w celu wyjaśnienia związku pomiędzy jedną zmienną ciągłą, a niezależnymi zmiennymi.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Wzór wielokrotnej regresji liniowej

2.1. Analiza danych

Poniżej przedstawiono udostępnione wskaźniki medyczne. Dane te zawierają najistotniejsze czynniki, które mogą wpływać na długość życia.

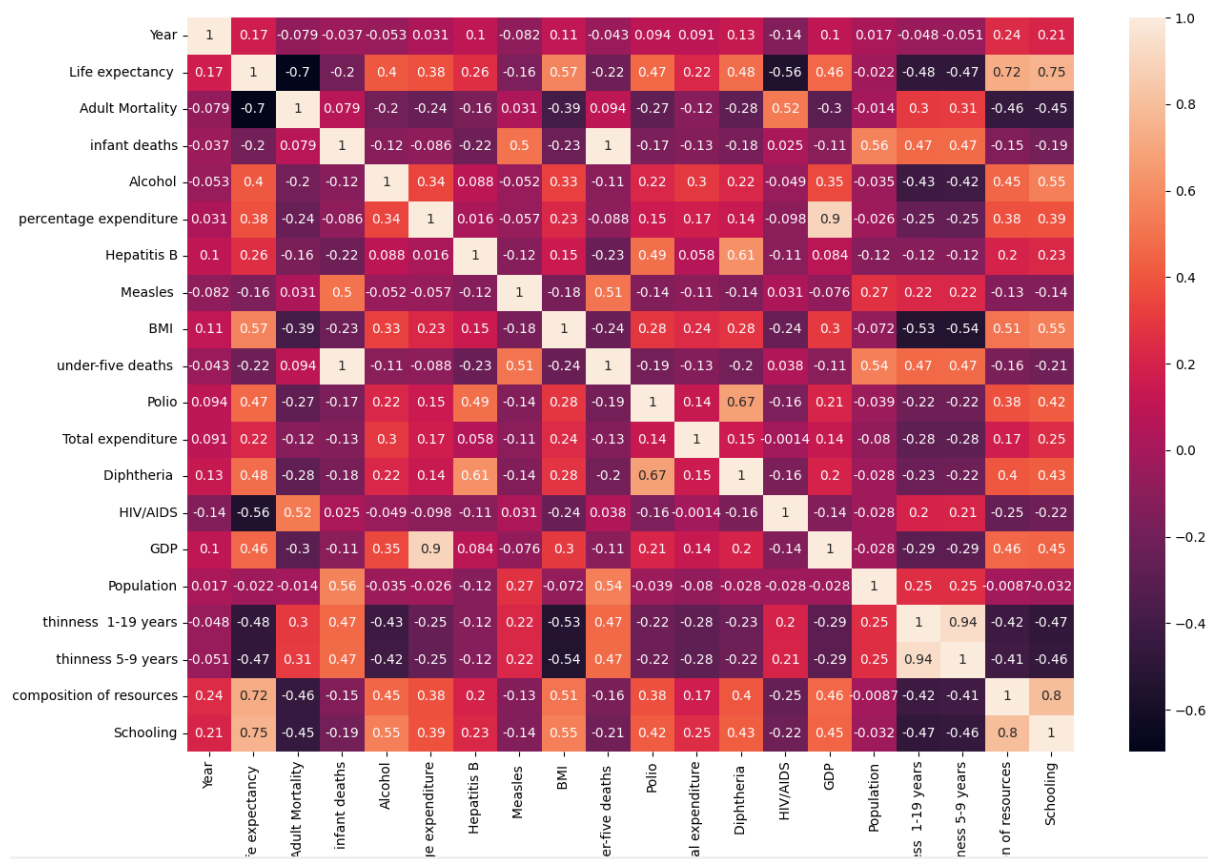
```
Country
Year
Status
Life expectancy
Adult Mortality
infant deaths
Alcohol
percentage expenditure
Hepatitis B
Measles
BMI
under-five deaths
Polio
Total expenditure
Diphtheria
HIV/AIDS
GDP
Population
thinness 1-19 years
thinness 5-9 years
Income composition of resources
Schooling
```

Badane wskaźniki

W celu poprawnej analizy dane tekstowe zastąpione zostały wartościami liczbowymi:

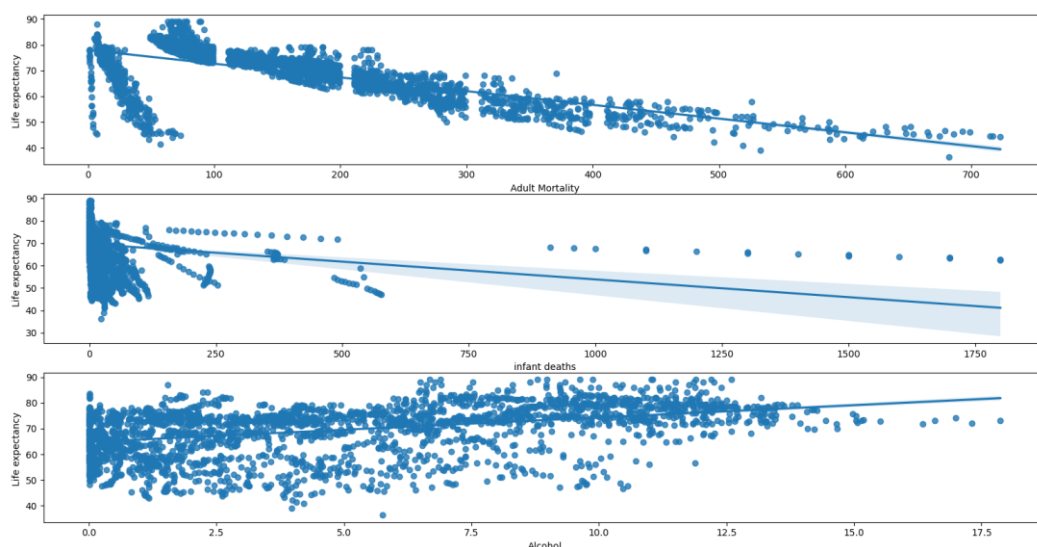
- Nazwa państwa – wartości od 0 do 192
- Stopień rozwoju danego kraju – 0 lub 1

Na poniższym rysunku przedstawione zostały zależności pomiędzy danymi zmiennymi.

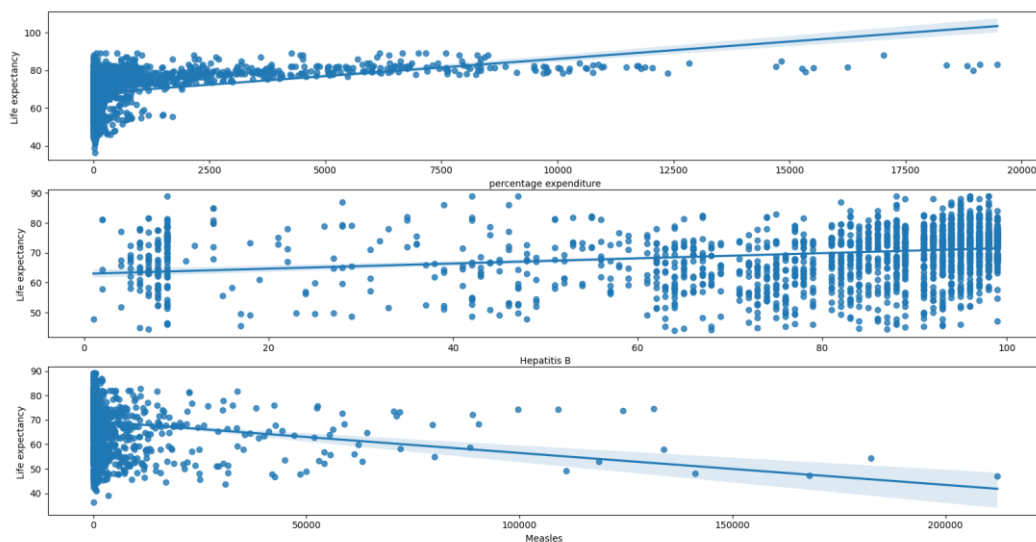


Rys. 1. Korelacja wskaźników przedstawiona za pomocą heatmapy

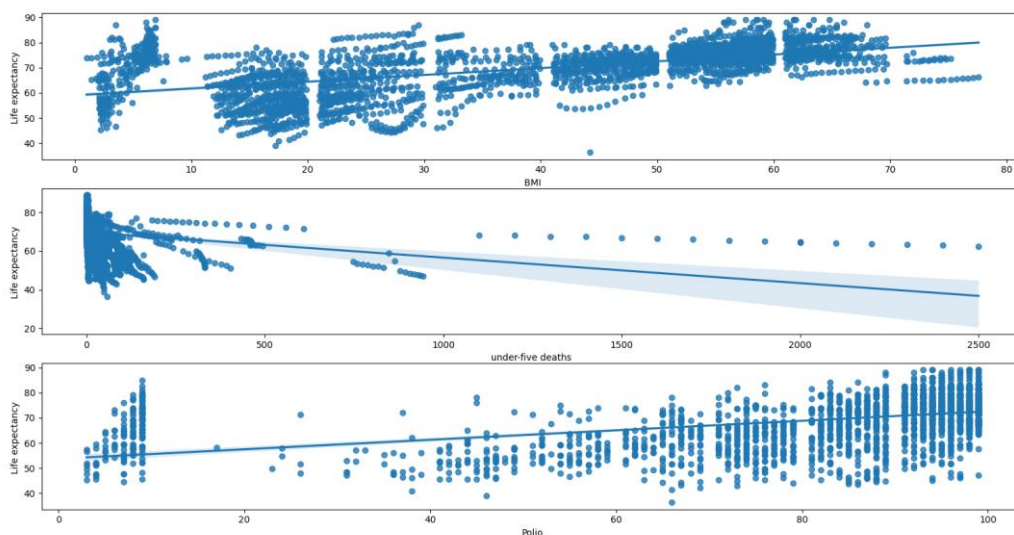
Poniżej przedstawione zostały wykresy długości życia pacjentów w funkcjach poszczególnych wskaźników medycznych. Niebieska linia oznacza wektor regresji liniowej otrzymany na podstawie użytych punktów.



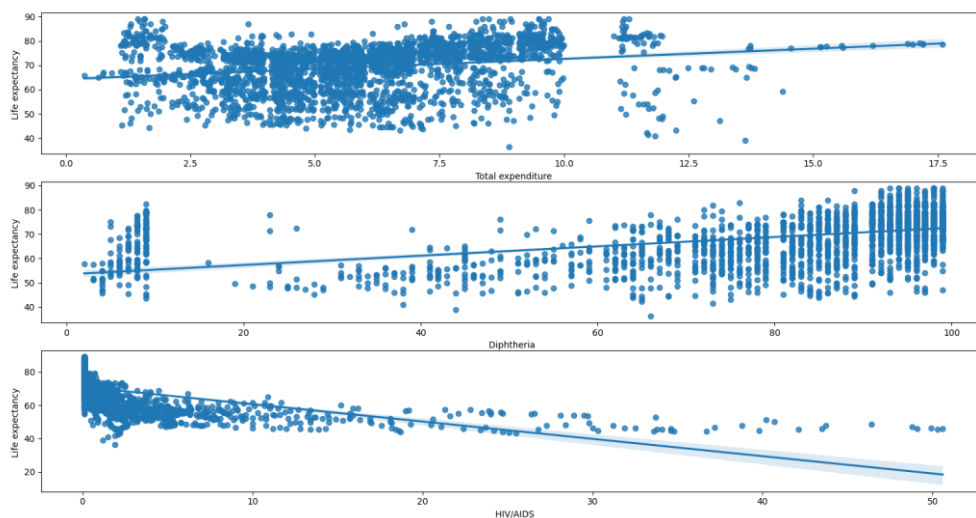
Rys. 2. Długość życia w zależności od: 1- śmiertelności dorosłych 2- śmierci dzieci 3- spożycia alkoholu



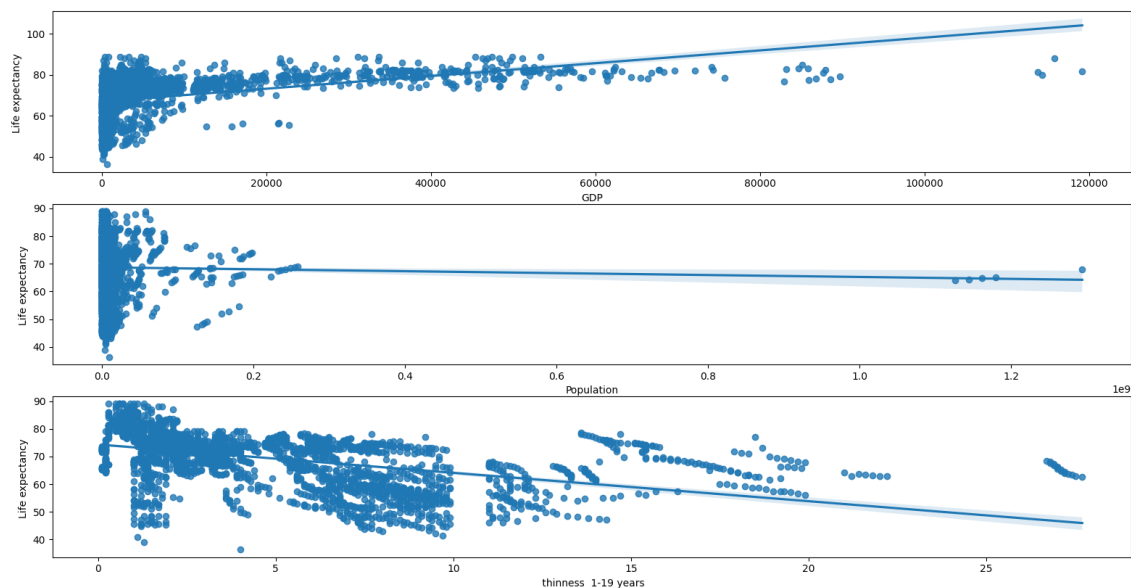
Rys. 3. Długość życia w zależności od: 1- procentowych wydatków 2- zapalenia wątroby typu B 3- odry



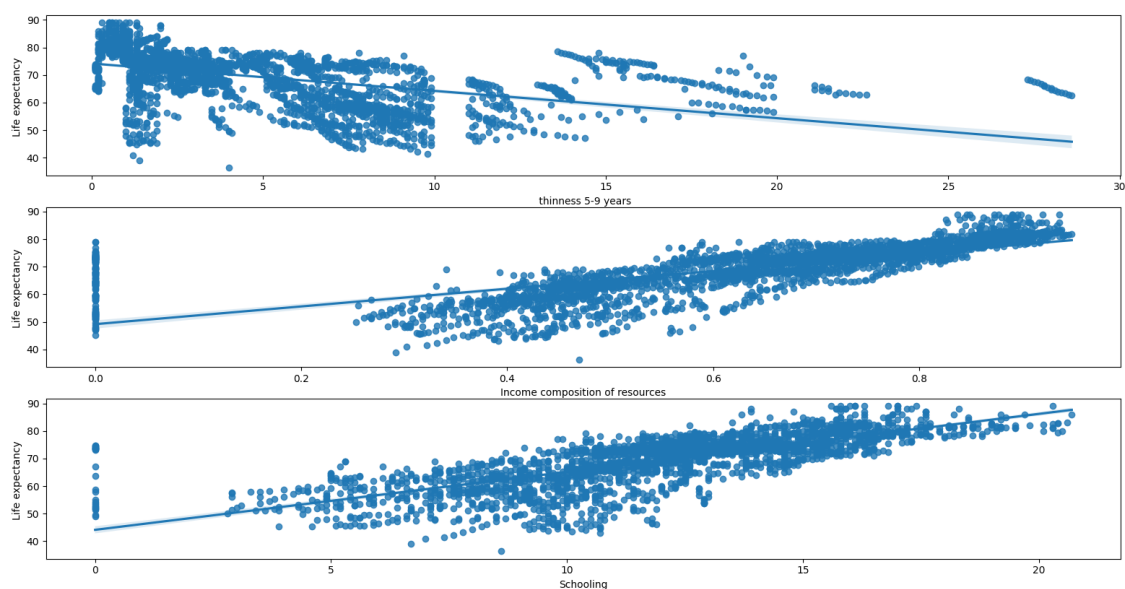
Rys. 4. Długość życia w zależności od: 1- BMI 2- śmierci poniżej 5 lat 3- polio



Rys. 5. Długość życia w zależności od: 1- wydatków 2- błonicy 3- HIV/ AIDS



Rys. 6. Długość życia w zależności od: 1- produktu krajowe brutto 2- populacji 3- niedowagi w wieku 1-19lat



Rys. 7. Długość życia w zależności od: 1- niedowagi w wieku 5-9lat 2 dystrybucji przychodów 3- poziomu szkolnictwa

Na podstawie tych wykresów możemy stwierdzić o wpływie i ważności danego czynnika na długość ludzkiego życia. W czynnikach, które mają znaczący wpływ można zobaczyć silny trend wzrostowy lub spadkowy. Pozostałe wskaźniki wykazują stały trend lub też w stosunkowo słaby trend wzrostowy/ spadkowy, co oznacza, że dane mają bardzo mały wpływ na długość życia pacjentów.

Czynniki z silnym trendem wzrostowym:

- dystrybucja przychodów
- poziom szkolnictwa

Czynniki z silnym trendem spadkowym:

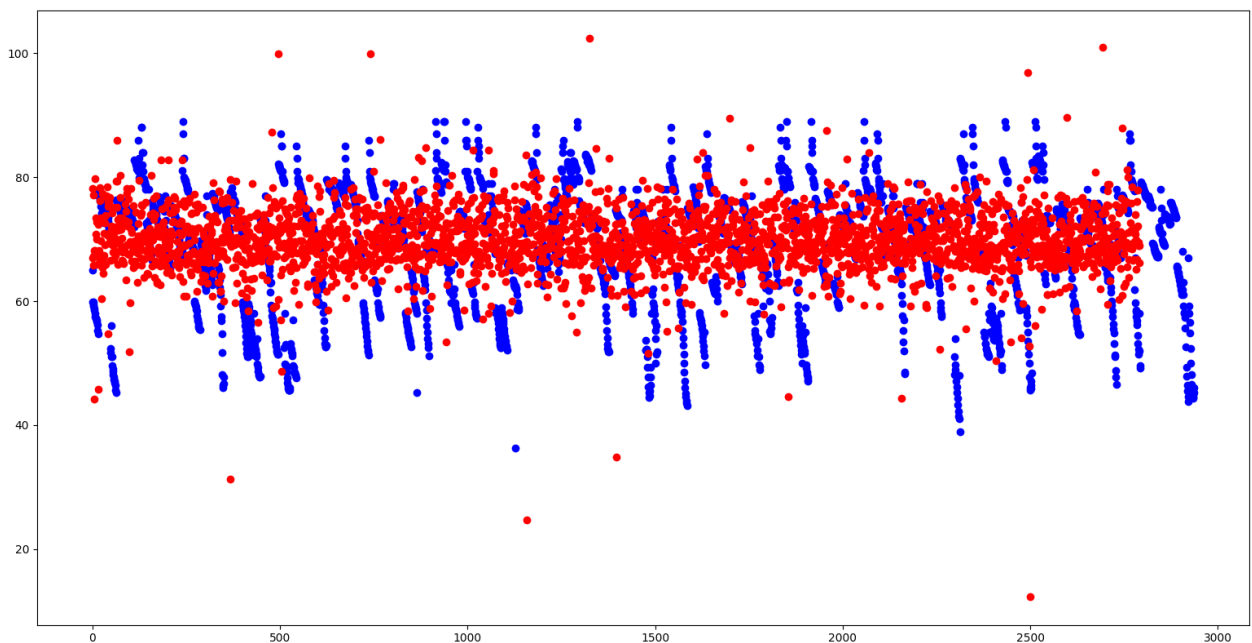
- śmiertelność dorosłych
- odra
- HIV/ AIDS
- niedowaga w wieku 1-19 lat oraz 5 -9 lat

Wymienione powyżej czynniki wpływają najmocniej na zwiększanie bądź też zmniejszanie długości życia pacjentów. Pozostałe wskaźniki przedstawione na rysunkach 2-7 mają mniejszy lub prawie żaden wpływ na długość życia.

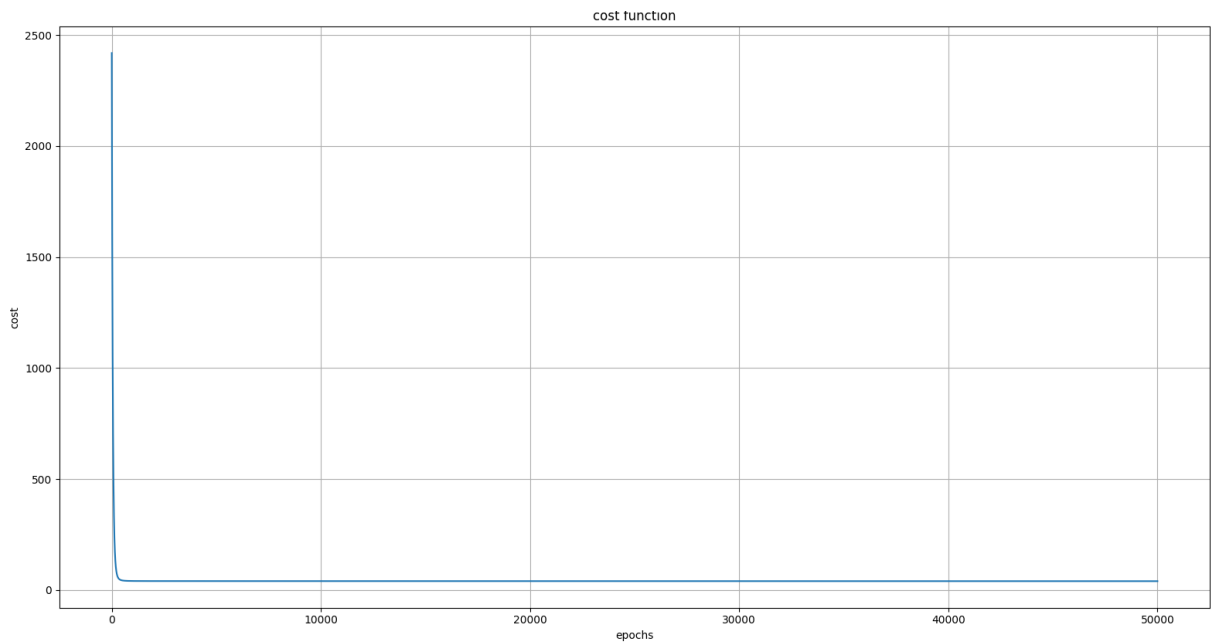
2.2. Otrzymane wyniki

Zbiór danych został podzielony na część przeznaczoną do uczenia się modelu oraz część testową. Dokładność działania algorytmu zależy od tego jak dużo danych zostanie przeznaczone na uczenie.

Na poniższym wykresie możemy zaobserwować porównanie wyników uzyskanych poprzez algorytm MLR (czerwone punkty) z wynikami rzeczywistymi (niebieskie punkty). Wykres przedstawia długość życia w funkcji pozostałych wskaźników medycznych.



Rys. 8. Porównanie długości życia z pliku oraz wyników otrzymanych przy pomocy modelu



Rys. 9. Postać funkcji kosztów w kolejnych iteracjach

```
Cost at epoch 0 is: 2464.8492007750874
Cost at epoch 5000 is: 10.637207012670874
Cost at epoch 10000 is: 8.264668842344049
Cost at epoch 15000 is: 7.814901027799637
Cost at epoch 20000 is: 7.652261333620664
Cost at epoch 25000 is: 7.580911019240708
Cost at epoch 30000 is: 7.544775078536691
Cost at epoch 35000 is: 7.523106195186292
Cost at epoch 40000 is: 7.5075820036250684
Cost at epoch 45000 is: 7.494764418272219
```

Rys. 10. Wartości funkcji kosztów w kolejnych iteracjach

3. Wnioski

Z wykresów zamieszczonych w punkcie 2.2. można stwierdzić poprawność działania algorytmu. Wykorzystuje on wszystkie dostępne wskaźniki medyczne i na ich podstawie prognozuje długość życia pacjentów, która zachowuje trend rzeczywistej. Funkcja kosztów przyjmuje niewielkie wartości już w początkowych iteracjach i utrzymuje swoją wartość przez kolejne.