



# Telematyka Medyczna

## System Predykcji Kosztów Leczenia Medycznego

Mateusz Miler 171577  
Sławomir Siwek 188175  
Łukasz Klein 172001

Katedra Inżynierii Biomedycznej, ETI



Fundusze  
Europejskie  
Polska Cyfrowa



Rzeczpospolita  
Polska

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego  
Program Operacyjny Polska Cyfrowa na lata 2014-2020.

Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.

Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.

# O projekcie

- Cel projektu - implementacja systemu predykcji kosztów leczenia medycznego
  - analiza stanu wiedzy w temacie projektu - artykuły naukowe, Internet
  - ocena trzech istniejących metod przewidywania kosztów opieki zdrowotnej
  - implementacja potencjalnie najlepszego rozwiązania wraz z wnioskami
- Docelowi użytkownicy
  - pacjenci - przewidywanie kosztów leczenia
  - lekarze - analiza czynników wpływających na rozwój kosztów leczenia
  - placówki medyczne - określenie najbardziej opłacalnego sprzętu medycznego i leków

# Funkcjonalności

- predykcja kosztów leczenia - pacjent
- wyodrębnienie czynników najbardziej wpływających na koszty opieki zdrowotnej - lekarz
- przewidywanie rozwoju kosztów opieki zdrowotnej w kolejnych latach - placówka medyczna
- kontrola szybko rosnących wydatków na opiekę zdrowotną - pacjent, placówka medyczna
- wykorzystanie wyników systemu wraz z wnioskami nad nimi do różnego rodzaju statystyk - pacjent, lekarz, placówka medyczna

# Analiza stanu wiedzy

1. M.A Morid, K. Kawamoto, T. Ault, J. Dorius, S. Abdelrahman - Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation, 16.04.2018  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977561/>
2. B. Panay, N. Baloian, J.A Piono, S. Penafiel, H. Sanson, N. Bersano - Predicting Health Care Costs Using Evidence Regression, 21.11.2019  
[https://mdpi-res.com/d\\_attachment/proceedings/proceedings-31-00074/article\\_deploy/proceedings-31-00074-v2.pdf](https://mdpi-res.com/d_attachment/proceedings/proceedings-31-00074/article_deploy/proceedings-31-00074-v2.pdf)
3. A.I Taloba, R.M Abd El-Aziz, H.M Alshanbari, A.H El-Bagoury - Estimation and Prediction of Hospitalization and Medical Care Costs Using Regression in Machine Learning  
<https://www.hindawi.com/journals/jhe/2022/7969220/>

# Artykuł nr 1. - opis

- Streszczenie kilku innych artykułów naukowych
- Dane - Uniwersytet Utah (2013-2016) 3.8 miliona danych medycznych i 780 tysięcy danych farmaceutycznych dla około 24 tysięcy różnych osób - dane demograficzne, dane kliniczne, diagnozy, procedury leczenia, koszty leczenia
- Wstępne przetwarzanie danych zgodnie z literaturą na temat bilansowania kosztów
- Modele - regresja liniowa, Lasso<sup>7</sup>, Ridge<sup>8</sup>, Elastic Net<sup>9</sup>, CART<sup>10</sup>, M5<sup>11</sup>, Losowy las<sup>12</sup>, Bagging<sup>13</sup>, Gradient Boosting<sup>13</sup>, SVM<sup>14</sup> i ANN<sup>15</sup>
- Wyniki - preferowane podejście do predykcji kosztów opieki medycznej powinno być w oparciu metody uczenia nadzorowanego, wybór najlepszego modelu może się różnić w zależności od wysokości analizowanych kosztów

# Artykuł nr 1. - rodzaje metod prognozowania kosztów

- metody oparte na regułach - wymagana duża wiedza dziedzinowa
- metody statystyczne - problem zbyt wielu danych zależnych współliniowo oraz duża podważalność wydajności
- metody nadzorowane - najlepsze podejście, które w głównej mierze było opisane w artykule

# Artykuł nr 1. - wyniki

- MAE - średni błąd modelu w przewidywaniu rzeczywistych wartości kosztów

$$MAE = \frac{\sum_i |a_i - p_i|}{n}$$

- MAPE - modyfikacja MAE

$$MAPE = \frac{\frac{\sum_i |a_i - p_i|}{n}}{\bar{a}}$$

- R<sup>2</sup> - korelacja Pearsona

$$R^2 = 1 - \frac{\sum_i (a_i - p_i)^2}{\sum_i (a_i - \bar{a})^2}$$

- Hit Radio - procent poprawnego kosztu

$$\text{Hit Ratio} = \frac{\text{Number of members with correct predicted basket}}{\text{Total number of members}}$$

# Artykuł nr 1. - wyniki

1. Gradient Boosting
2. ANN
3. Ridge
4. SVM
5. ElasticNet
6. Lasso
7. M5
8. Regresja Liniowa
9. Las losowy
10. Bagging
11. CART



# Artykuł nr 2. - podsumowanie

- Artykuł bazujący na Japońskiej dokumentacji medycznej
- Dane - dokumentacja medyczna szpitalu Tsuyama Chuo (2016-2017)
  - informacja o pacjencie (płeć, wiek itd.), informacja o objawach, leczeniu, kosztach
- Modele: IEVREG, Gradient Boosting, ANN
- Wyniki:
  - a. IEVREG - autorska metoda regresji
  - b. GB
  - c. ANN
- Ponownie zastosowano te same modele (poza pierwszym) i udało im się stworzyć własny model, który przebił standardowe rozwiązania

# Artykuł nr 3. - podsumowanie

- Dotyczy aspektu kosztów leczenia otyłości na podstawie wskaźników BMI - powiązane z inną pracą naukową
- Podejście statystyczne wraz z analizą biznesową
- Dane: informacje od ponad 7.5 tysiąca pacjentów z CHF (frank szwajcarski) hospicjalizowanych w Szwecji - dane personalne, wskaźniki BMI, koszty
- Modele: Regresja liniowa, naiwny klasyfikator Bayesa, lasy losowe
- Wyniki: Regresja liniowa, lasy losowe, naiwny klasyfikator Bayesa
- Rozwiązania bardzo podobne do Artykułu 2. - obie prace nazwały swoje metody Regresji liniowej autorskimi

# Podsumowanie artykułów naukowych







- Istnieje masa metod pozwalających na predykcję kosztów opieki medycznej
- Istnieją trzy podejścia do problemu: metody oparte na regułach, metody oparte na statystyce, metody uczenie nadzorowanego
- Najczęściej pojawiające się rozwiązania to: regresja liniowa, Gradient Boosting oraz ANN
- Predykcja kosztów leczenia może być wykorzystywana na wielu płaszczyznach
- Wybór datasetu ma równie duże znaczenie, co wybór metody
- Najlepsze rozwiązania operowały w okolicach dziewięćdziesięciu kilku procent poprawnych predykcji w zakresie ogólnych kosztów opieki medycznej

# Schemat działania aplikacji



# Dostępność zbiorów danych

Tylko dwa unikalne zbiory danych - oba dotyczą *ubezpieczenia!*

 <b>Medical Cost Personal Datasets</b> by PAVAN KUMAR D. a year ago • 16 kB • ^ 13 <a href="#">Medical Cost Personal Datasets</a>	 <b>Medical Cost Personal Datasets</b> by dwiuzila a year ago • 161 kB • ^ 0 <a href="#">Medical Cost Personal Datasets</a>	 <b>Medical Cost Personal Datasets</b> by Miri Choi 4 years ago • 16 kB • ^ 1717 <a href="#">Medical Cost Personal Datasets</a>
 <b>Medical Cost Personal Dataset</b> by Abdel Homi 2 years ago • 16 kB • ^ 5 <a href="#">Medical Cost Personal Dataset</a>	 <b>medical insurance cost stat analysis</b> by Bishnusaran Sahu a year ago • 16 kB • ^ 0 <a href="#">medical insurance cost stat analysis</a>	 <b>Medical Insurance Premium Prediction</b> by Tejashvi 8 months ago • 7 kB • ^ 63 <a href="#">Medical Insurance Premium Prediction</a>

# Medical Cost Personal Dataset

- udostępniony w ramach książki “Machine Learning with R”
- wymaga przetwarzania wstępnego
- 6 cech
- 1339 rekordów
- do oceny kosztów leczenia  
najczęściej wykorzystywane są:
  - regresja liniowa,
  - regresja wielomianowa,
  - Decision Tree Regressor,
  - Random Forest Regressor.

age	sex	bmi	children	smoker	region	expenses
19	female	27.9	0	yes	southwest	16884.92
18	male	33.8	1	no	southeast	1725.55
28	male	33	3	no	southeast	4449.46
33	male	22.7	0	no	northwest	21984.47
32	male	28.9	0	no	northwest	3866.86
31	female	25.7	0	no	southeast	3756.62
46	female	33.4	1	no	southeast	8240.59
37	female	27.7	3	no	northwest	7281.51
37	male	29.8	2	no	northeast	6406.41
60	female	25.8	0	no	northwest	28923.14

# Plan prac

- opracowanie interfejsu użytkownika aplikacji
- implementacja API obsługującego zapytania z formularza aplikacji
- oczyszczenie zbioru danych,
- trening kilku modeli regresji, wybór najlepszego,
- scalenie wszystkich modułów aplikacji,
- test działania aplikacji,
- opracowanie dokumentacji oraz demo



# Dziękujemy za uwagę



**Fundusze  
Europejskie**  
Polska Cyfrowa



**Rzeczpospolita  
Polska**

**Unia Europejska**  
Europejski Fundusz  
Rozwoju Regionalnego



Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego

Program Operacyjny Polska Cyfrowa na lata 2014-2020.

Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.

Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.