



Ministerstwo Zdrowia

Departament Analiz i Strategii Ministerstwa Zdrowia (DAiS MZ) zajmuje się analizą dużych zbiorów danych na potrzeby optymalizacji działania służby zdrowia w Polsce. W ramach współpracy z SGH pracownicy departamentu przygotowali 3 zadania w oparciu o dane płatnika (NFZ).

Przekazane dane dotyczą pacjentów udarowych w 2016 roku oraz 2-letniej historii ich chorób. Dane odnoszą się do ogółu populacji Polski. Przedmiotowe zadania są jednymi z przykładowych wyzwań, które stanowią rdzeń aktywności analitycznej DAiS MZ. W oparciu o sugestie ekspertów medycznych oraz specjalistów z zakresu zdrowia publicznego formułujemy kluczowe benchmarki służące ocenie jakości leczenia i diagnostyki najbardziej istotnych rozpoznań chorobowych z punktu kondycji zdrowia populacji. Biorąc pod uwagę wartość budżetu NFZ (na poziomie sumy kapitalizacji kilku wiodących spółek z indeksu WIG20) oraz procent populacji korzystających z usług finansowanych przez NFZ co najmniej raz w ciągu roku, optymalna alokacja środków powierzonych NFZ wydaje się być kluczowym zadaniem dla całego systemu ochrony zdrowia.

Treść zadań brzmi następująco:

- 1) Przygotuj model predykcyjny w oparciu o dane dotyczące pacjentów z udarem w 2016 r. (np. analiza przeżycia, śmiertelność pacjentów).
- 2) Przygotuj model optymalizujący rozmieszczenia oddziałów udarowych w Polsce (minimalizacja czasu dojazdu pacjenta, polepszenie jakości opieki nad pacjentem). Założenia modelu:
 - a) pacjent powinien trafić na oddział udarowy w ciągu 30 min od zdarzenia,
 - b) przyjmuje się, że na 200 tys. ludności powinien przypadać co najmniej jeden oddział udarowy,
 - c) każdy oddział powinien przyjmować co najmniej 300 pacjentów rocznie.
- 3) Przygotuj interesujące wizualizacje na bazie dostępnych danych.

Dane udostępnione przez DAiS MZ przekazane zostały w 2 plikach `dane_udary.csv` (ok. 134 tys. pacjentów, 150 tys. rekordów, 10 zmiennych) oraz `historia_pacjenta_udarowego.csv` (350 tys. rekordów, 5 zmiennych). Te tabele można połączyć przyjmując za klucz należy przyjąć zmienne `ID_PACJ` oraz `DATA_UDARU`. Format danych to UTF-8. W analizie demograficznej należy posługiwać się danymi GUS (<https://bdl.stat.gov.pl/BDL/start>).

Efektom pracy powinien być raport, składający się z **maksymalnie 20 stron**.

Przykłady

Przykładowa analiza pacjenta:

- Przypadek 1:
 - Pacjent o `ID_PACJENTA 01714329E56B6B7D233DF613922FA071`.
 - Hospitalizowany z powodu udaru 25 października 2016 r., wypisano go ze szpitala 2 listopada 2016 r. Był nim mężczyzna w wieku 81 lat. Kod `TERYT_PACJENTA` wskazuje, że był on zameldowany w pow. m. Legnicy (woj. dolnośląskie), w tej samej miejscowości się leczył. Pacjent nie został poddany procedurze trombolizy. W ciągu 180 dni od hospitalizacji nie odnotowano zgonu pacjenta ani nie był ponownie hospitalizowany z powodu udaru w ciągu 90 dni.

```
dane_udary[ID_PACJ == "01714329E56B6B7D233DF613922FA071", -3]
```

- o Analiza historii pacjenta wskazuje, że 397 dni przed udarem chorował na nadciśnienie tętnicze, następnie (158 dni przed udarem) chorował na objawy i cechy chorobowe dotyczące układu pokarmowego i jamy brzusznej. Natomiast 40 dni przed udarem (15 września 2016 r.) cierpiał na zaburzenia soczewki.

```
historia_pacjenta_udarowego[ID_PACJ == "01714329E56B6B7D233DF613922FA071", 4:5]
```

- Przypadek 2:

- o Pacjent o ID 1DA4C2A7C427DA6AF24D01D8D5AE1A28.
- o Była to 84-letnia kobieta hospitalizowana z powodu udaru 27 stycznia 2016 r., ze szpitala została wypisana 8 lutego 2016 roku. Była zameldowana na warszawskim Mokotowie, leczyla się w szpitalu mającym siedzibę na terenie m.st. Warszawy (nieznana dzielnica). Ponieważ m.st. Warszawa jest miastem na prawach powiatu, można przyjąć, że była zameldowana i leczyla się na terenie tego samego powiatu (4-znakowy TERYT_PACJ = 4-znakowy TERYT_KOMORKI). Pomimo poddania się procedurze trombolizy kobieta miała ponowny udar i w jego następstwie umarła po 32 dniach.

```
dane_udary[ID_PACJ == "1DA4C2A7C427DA6AF24D01D8D5AE1A28", -3]
```

- o Analiza historii pacjentki wskazuje, że 504 dni przed udarem chorowała na objawy i cechy chorobowe ogólne. Następnie, 307 dni przed udarem, chorowała na zaburzenia soczewki. Natomiast 146 dni przed udarem odnotowano u niej choroby wątroby, 2 dni później choroby pęcherzyka żółciowego, dróg żółciowych oraz trzustki.

```
historia_pacjenta_udarowego[ID_PACJ == "1DA4C2A7C427DA6AF24D01D8D5AE1A28", 4:5]
```

- Przypadek 3:

- o Pacjent o ID_PACJ 571ED9191B00FE586BA624A039AD6549.
- o Był to 40-letni mężczyzna, którego hospitalizowano z powodu udaru 21 listopada 2016 roku, wypisany ze szpitala 1 grudnia 2016 r. Był zameldowany w gminie Końskie, leczyl się w mieście będącym siedzibą tejże gminy miejsko-wiejskiej (woj. świętokrzyskie). Podczas hospitalizacji został poddany procedurze trombolizy. Nie odnotowano ponownego udaru do 90 dni ani zgonu do 180 dni po hospitalizacji.

```
dane_udary[ID_PACJ == "571ED9191B00FE586BA624A039AD6549", -3]
```

- o Analiza historii pacjenta wskazuje, że 253 dni przed udarem przeżył wypadek. Następnie chorował na chorobę tkanek miękkich (91 dni przed udarem) oraz na kamice moczową (79 dni przed udarem).

```
historia_pacjenta_udarowego[ID_PACJ == "571ED9191B00FE586BA624A039AD6549", 4:5]
```

Opis zmiennych

Tabela 1: Opis zbioru historia_pacjenta_udarowego.csv

zmienna	typ danych		opis
ID_PACJ	character		Unikalne ID Pacjenta
DATA_ROZPOZNANIA	date	YYYY-MM-DD	Data wystąpienia choroby
DATA_UDARU	date	YYYY-MM-DD	Data wystąpienia udaru
DNI_PRZED_UDAREM	numeric		Różnica między data udaru a poprzedzającym go rozpoznaniem
KATEGORIA_ROZPOZNANIA	character		Typ rozpoznania

Tabela 2: Opis zmiennych zbioru `dane_udary.csv`

zmienna	typ danych		opis
DATA_UDARU	date	YYYY-MM-DD	Data wystąpienia udaru
DATA_WYPISU	date	YYYY-MM-DD	Data wypisu ze szpitala
ID_PACJ	character		Unikalne ID Pacjenta
PLEC	nominal	K/M	Płeć pacjenta: K – kobieta, M – mężczyzna
TERYT_KOMORKI	character	siedmioznakowy	Kod TERYT ^a Świadczeniodawcy (gmina udzielenia świadczenia)
TERYT_PACJ	character	siedmioznakowy	Kod TERYT ^a Pacjenta (gmina zamieszkania pacjenta)
WIEK	integer		Wiek pacjenta
CZAS_DO_ZGONU	numeric		Po ilu dniach wystąpił zgon pacjenta: –1 – zgon nieodnotowany w przeciągu 180 dni, 0 – zgon wystąpił w dniu udaru, x – zgon nastąpił x dni po udarze
PONOWNY_UDAR_90_DNI	logical	true/false	Czy wystąpił ponowny udar w ciągu 90 dni od daty tego udaru?
TROMBOLIZA	logical	1/0	Procedura trombolizy świadczy o wysokiej jakości leczenia pacjenta: 1 – pacjentowi wykonano procedurę trombolizy, 0 – pacjentowi nie wykonano procedury trombolizy

^a <https://pl.wikipedia.org/wiki/TERYT>, <https://pl.wikipedia.org/wiki/TERC>