

## WYKORZYSTANIE OPROGRAMOWANIA CLEMENTINE

# *do wykrywania przypadków nadużyć i niegospodarności w Śląskim Oddziale Wojewódzkim NFZ*

PROJEKT  
PILOTAŻOWY

### 1. Sytuacja

Narodowy Fundusz Zdrowia jest instytucją zarządzającą ogromnymi środkami finansowymi, które sięgają 50 mld PLN rocznie. Zarządzanie tak dużą kwotą stanowi ogromną odpowiedzialność w zakresie właściwej alokacji środków, zapewnienia realizacji wszystkich istotnych dla społeczeństwa świadczeń medycznych i zagwarantowania racjonalności podejmowanych wydatków.

Skala budżetu NFZ, ogromna ilość realizowanych świadczeń zdrowotnych, refundacji leków stanowi naturalny obszar dla występowania niegospodarności, nadużyć i oszustw. Wyludzone z NFZ środki finansowe stanowią szczególną stratę, jeżeli spojrzymy przez pryzmat skutecznego zarządzania zasobami, gdyż środki zmarnotrawione nie zostaną wykorzystane na inne, ważne społecznie cele.

Doświadczenia pracowników NFZ, przykłady faktycznie wykrytych nadużyć oraz doświadczenia innych krajów dowodzą, że zjawisko wyludzenia pieniędzy z systemu ubezpieczeń zdrowotnych nie jest marginalne i każdego roku generuje trudne do oszacowania straty. Można pokusić się o założenie, że 10-20% środków jest corocznie marnotrawionych, co stanowi ogromną dla budżetu Państwa kwotę 5-10 mld PLN rocznie.

Z pewnością zatem działania mające na celu wykrywanie nadużyć i niegospodarności oraz analiza systemu pod kątem zwiększenia racjonalności wydatków są jednym z głównych zadań realizowanych przez NFZ. Działania te stanowią jeden z filarów realizacji misji NFZ, którą jest „(...) zapewnienie uprawnionym osobom dostępu do opieki zdrowotnej na najwyższym możliwym poziomie poprzez efektywne gospodarowanie środkami finansowymi powierzonymi NFZ na realizację tego celu”.

Aktualnie wykrywaniem nadużyć zajmują się wyspecjalizowane komórki kontroli funkcjonujące w ramach NFZ, jednak skuteczność podejmowanych przez nie działań jest ograniczona przez ogromną liczbę podmiotów do skontrolowania.

Konieczne jest zatem zastosowanie zaawansowanych narzędzi analizy danych, które spośród ogromu gromadzonych danych pozwolą na bardziej trafne typowanie podmiotów do kontroli. NFZ jest świadomy konieczności wdrożenia takiego narzędzia, co zostało zapisane w dokumencie „Strategia wykorzystania zasobów informacyjnych przez NFZ oraz kierunki rozwoju systemu informatycznego NFZ na lata 2007-2010”, która określa, że będą podejmowane działania mające na celu budowę Systemu Fraud Detection w każdym z regionalnych oddziałów NFZ.

## 2. Rozwiązanie

Zwiększenie skuteczności realizowanych przez NFZ kontroli nie jest zadaniem łatwym i wymaga zastosowania odpowiednich narzędzi analizy danych. SPSS POLSKA jest dostawcą profesjonalnego oprogramowania, które na całym świecie znalazło zastosowanie w realizacji różnych zadań analitycznych. Instytucje gromadzące duże wolumeny danych korzystają z oprogramowania klasy data mining, które pozwala na optymalną i efektywną analizę danych, raportowanie czy budowanie modeli predykcyjnych.

Rozwiązania data mining pomagają firmom komercyjnym pozyskiwać nowych klientów, optymalizować kampanie marketingowe, zapobiegać odchodzeniu klientów do konkurencji, szacować ryzyko spłaty kredytu itp. W rezultacie firmy mogą zwiększyć zyski i równocześnie ograniczyć koszty operacyjne. Dlatego też użytkownikami narzędzi data mining są banki, firmy ubezpieczeniowe, telekomunikacyjne, usługowe, czy też windykacyjne.

Data mining oferuje również szerokie spektrum zastosowań dla administracji publicznej, która jest właścicielem ogromnej ilości danych. Patrząc tylko przez pryzmat Polski, kopalnie danych to choćby: NFZ, ZUS, ARIMR, MINISTERSTWO FINANSÓW, MINISTERSTWO PRACY I POLITYKI SPOŁECZNEJ, MINISTERSTWO GOSPODARKI, POLICJA i inne.

Klasycznym celem administracji publicznej jest dbałość o efektywność i zasadność wydawania środków publicznych. Dlatego data mining został zaprzęgnięty do wykrywania nadużyć na całym świecie, w tym również w Polsce.

Narzędziem data mining, które SPSS z powodzeniem wdraża w różnych instytucjach w celu poszukiwania nadużyć i oszustw jest CLEMENTINE. Firma SPSS POLSKA zaproponowała Śląskiemu Oddziałowi NFZ realizację pięciodniowego projektu pilotażowego, którego celem miało być przetestowanie przydatności zastosowania specjalistycznego oprogramowania w celu zwiększenia wykrywania nadużyć i skuteczności realizowanych kontroli.

### 2.1. OGRANICZANIE NADUŻYĆ DZIĘKI DRĄŻENIU DANYCH

Oddziały NFZ aktywnie usiłują zredukować oszustwa i nadużycia w obrocie lekami i ich refundacji. Jednak oszuści często posługują się wymyślnymi i skomplikowanymi schematami, by uzyskać nienależną refundację leków. Ze względu na ograniczenia występujące w zakresie ilości możliwych do zrealizowania kontroli instytucje odpowiedzialne za prowadzenie działań kontrolnych mają duże problemy z walką z oszustwami i nadużyciami w refundacji leków.

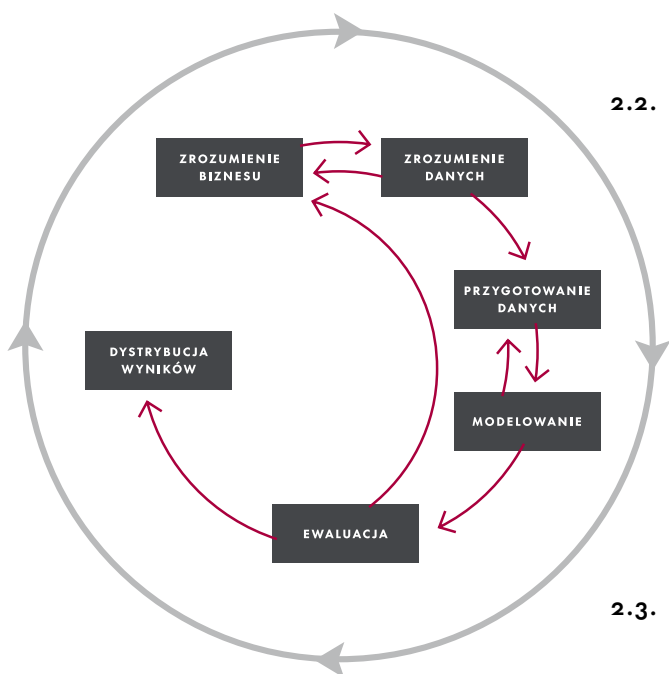
Zespoły kontrolujące zazwyczaj posługują się intuicją i zdobytym doświadczeniem, aby skonstruować profile lekarzy i aptek o skłonności do dokonywania nadużyć. Jednak te nienaukowe sposoby typowania lekarzy i aptek do kontroli powodują, że kontrolerzy poświęcają dużo czasu na sprawdzanie niekoniecznie tych podmiotów, które dokonują największych nadużyć.

Dlatego też SPSS zaoferował proces drążenia danych, jako sposób na skuteczne wykrywanie nadużyć i efektywne prowadzenie działań kontrolnych. Coraz większa liczba instytucji zarządzających wydatkowaniem funduszy zdrowotnych na świecie wprowadza modele drążenia danych dla celów typowania do kontroli podmiotów podejrzanych o nadużycia. Kontrolerom dostarczane są informacje o podmiotach, którzy z wysokim prawdopodobieństwem mogą być źródłem nadużyć. Algorytmy analityczne uzyskane w ramach drążenia danych łączone są z wiedzą i doświadczeniem kontrolerów, po to, aby informacje zawarte w danych mogły być wykorzystane do wykrywania nowych przypadków nadużyć.

Drążenie danych pozwala instytucjom zarządzającym wydatkowaniem funduszy zdrowotnych na:

- określenie kluczowych czynników umożliwiających wskazanie, które z refundacji są wypełnione nieprawidłowo,
- odkrycie reguł, które mogą zostać użyte przez kontrolerów do typowania refundacji, które najprawdopodobniej są związane z nadużyciami,
- zapewnienie kontrolerom odpowiedniej wiedzy umożliwiającej efektywne typowanie refundacji, będących prawdopodobnie oszustwami, a będących równocześnie źródłem dużych strat dla budżetu państwa.

W rezultacie przeprowadzenia procesu drążenia danych, instytucje kontrolne skupiają się bardziej na działaniach pozwalających na skuteczne powiększanie odzyskiwanych środków finansowych, zwiększając przy tym swój budżet na finansowanie innych programów.



Rysunek 1.

## 2.2. ODKRYWANIE WZORÓW OSZUSTW I NADUŻYĆ

Proces drążenia danych rozpoczyna się od etapu gromadzenia i zrozumienia dostępnych danych. W tym etapie weryfikowane jest czy dane nadają się do analizy, czy też potrzebne jest dostarczenie dodatkowych informacji, które mogą poprawić rezultaty prowadzonych analiz. Kolejnym etapem drążenia danych jest budowanie wzorów, tj. wyszukiwanie reguł postępowania podmiotów dokonujących nadużyć, a w dalszej kolejności budowa modeli, które potrafią te wzorce odnajdywać w danych i wskazać przypadki występowania nadużyć. Powyższe etapy drążenia danych stanowią klucz do właściwego połączenia wiedzy płynącej z danych i doświadczenia posiadanego przez kontrolerów. Bazując na tych dwóch elementach można w skuteczny sposób odnajdywać czynniki wskazujące na wysokie prawdopodobieństwo wystąpienia nadużycia.

## 2.3. BUDOWANIE MODELI ANALITYCZNYCH DO WYKRYWANIA NADUŻYĆ

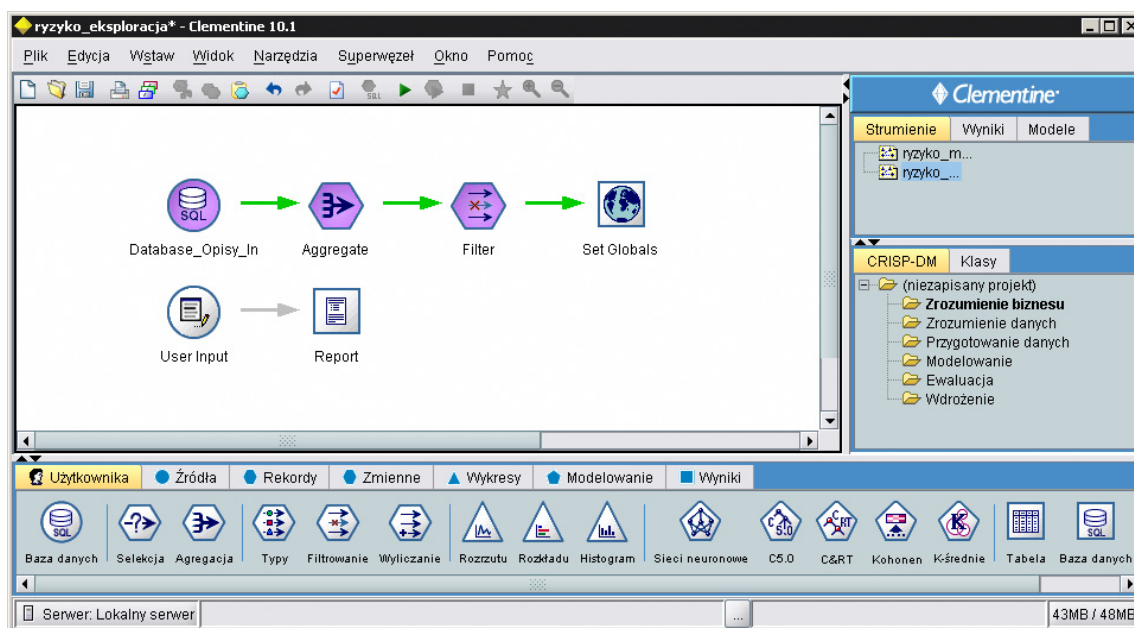
Zadaniem zespołów kontrolnych jest odnajdywanie przypadków refundacji, które mogą być źródłem nadużyć. Refundacje leków dla których w ramach przeprowadzonej kontroli stwierdzono nadużycia stanowią wzór do odzyskiwania pieniędzy w przyszłości. Z wykorzystaniem danych historycznych, w procesie drążenia danych możemy szczegółowo zweryfikować wszystkie refundacje zrealizowane w danym roku pod kątem możliwości wystąpienia w nich nadużyć.

Istniejące dane analizowane są z wykorzystaniem różnych technik analitycznych, począwszy od prostych raportów, a skończywszy na zaawansowanych metodach analizy zachowań lekarzy i aptek. W chwili, gdy gotowy jest model, należy upewnić się, czy działa on wystarczająco skutecznie z założonym poziomem ufności. Dla osiągnięcia tego celu można wykorzystać dodatkowe techniki analityczne pozwalające określić, które z algorytmów są efektywne, a które nie i w jakim stopniu. Ponieważ dane, na których pracują modele ulegają ciągłym zmianom, modele powinny być okresowo sprawdzane i aktualizowane. W ten sposób będą one dostarczać zawsze aktualnych rekomendacji.

## 2.4. WYKORZYSTYWANIE REZULTATÓW ANALIZ

W wyniku realizacji gotowych modeli analitycznych generowane są wyniki, które powinny zostać przekazane osobom decyzyjnym, w celu ich wykorzystania dla ograniczenia pojawiających się przypadków nadużyć. Wykorzystanie wyników oznacza również zintegro-

wanie modeli z procedurami działania instytucji, by tym łatwiej korzystać z wyników analiz w ramach codziennej pracy. Zweryfikowane pod względem skuteczności modele analityczne lub ich wyniki przekazywane są do komórki dokonującej kontroli, która w następnym kroku wykorzystuje je do klasyfikacji zgromadzonych w bazie przypadków refundacji. W ten sposób system analityczny oznacza te przypadki, które charakteryzują się największym prawdopodobieństwem wystąpienia nadużycia.



Rysunek 2. Zidentyfikowane rozliczenia powinny zostać uszeregowane ze względu na dwa kryteria: prawdopodobieństwo dokonania oszustwa i wysokość przewidywanej sumy oszustwa, którą można odzyskać w wyniku korekty. Po skonstruowaniu takiego rankingu może on zostać udostępniony w formie zbiorczej dla kadry zarządzającej, bądź szczegółowej dla kontrolerów, z wykorzystaniem raportów pisemnych, poczty elektronicznej czy intranetu.

Odpowiednia dystrybucja wyników wewnątrz instytucji umożliwia dostarczenie odpowiednim osobom i w odpowiednim czasie informacji niezbędnych dla przeprowadzenia kontroli. Instytucje dokonujące kontroli wspólnie z lokalnymi urzędami i biurami, wykorzystując drążenie danych w połączeniu ze zintegrowanymi systemami gromadzenia i dystrybucji danych dla skutecznego zwalczania występowania nadużyć. Dzięki stałemu przepływowi aktualnych informacji kontrolerzy mogą pracować bardziej efektywnie, kierując swoje wysiłki na zweryfikowanie tych przypadków, które wiążą się z nadużyciami i których korekta przyniesie największe zwroty pieniędzy lub w największym stopniu wpłynie na zmniejszenie strat finansowych.

### 3. Realizacja projektu

Projekt był realizowany w pierwszej połowie lutego 2008 r. i przeznaczono na niego 5 dni roboczych. Miejsmem realizacji była siedziba Oddziału Śląskiego NFZ, a zespół projektowy stanowiło: 3 pracowników Wydziału Gospodarki Lekami oraz 2 konsultantów SPSS POLSKA.

Ze względu na ograniczoną ilość czasu oraz brak możliwości wykorzystania architektury klient-serwer jak również czysto pilotażowy charakter przedsięwzięcia, zdecydowano się

*Funkcjonalności Clementine pozwalają na elastyczną i efektywną pracę z danymi zgromadzonymi w bazie danych. Byliśmy pozytywnie zaskoczeniu możliwością optymalizacji zapytań sql, która zdecydowanie upraszcza pracę i pozwala zaoszczędzić ogromną ilość czasu spędzaną przez nas wcześniej nad konstrukcją zapytań SQL. Możliwości analitycznego przetwarzania danych oferowane przez Clementine to z kolei funkcjonalność, jakiej nie byliśmy w stanie uzyskać w ramach wcześniej wykorzystywanego oprogramowania. Szczególnie ciekawe było tutaj zastosowanie technik modelowania, które pozwoliły odkryć wielowymiarowe zależności między zmiennymi*

”

Renata Merta  
Kierownik Działu  
Monitorowania i Analiz  
Wydział Gospodarki Lekami  
ŚOW NFZ

pracować jedynie na wybranych obszarach danych. W tym celu wyodrębnione zostały następujące zbiory danych:

- refundacja leków narkotycznych w latach 2006–2007 (107 627 rekordów);
- refundacja leków insulinowych za rok 2007 (1 153 991 rekordów);
- wielokrotne wystąpienia tych samych recept w roku 2007, wszystkie refundowane leki (33 037 rekordów).

W wyniku przeprowadzenia projektu pilotażowego zrealizowano poniżej opisane analizy.

### **3.1. WYODRĘBNIE ZBIORÓW ANALITYCZNYCH**

Wykorzystując funkcjonalność systemu CLEMENTINE dokonano selekcji i połączenia udostępnionych przez NFZ zbiorów poprzez:

- połączenie się z Oddziałową Bazą Danych (Oracle) z wykorzystaniem sterowników ODBC;
- selekcję wybranych grup leków za dany okres;
- połączenie istotnych informacji z trzech tabel bazy danych w jeden zbiór analityczny.

Dzięki funkcjonalności CLEMENTINE wygenerowano odpowiednie zapytania SQL i zlecono bazie danych ich realizację bez konieczności samodzielnego tworzenia zapytań SQL. Znaczna część operacji mocno obciążających zasoby komputerów lokalnych została przekazana na serwer bazodanowy (SQL pushback) co zdecydowanie przyspieszyło czas realizowanych operacji.

### **3.2. PRZYGOTOWANIE DANYCH DO ANALIZ**

W związku z założonym krótkim okresem realizacji projektu nie wykonano fazy szczegółowego audytu danych. Przeprowadzono jednak analizę duplikatów, które same w sobie okazały się potencjalnym źródłem wiedzy o nadużyciach (rozdział 3.4 „Zrealizowane analizy – duble recept w roku 2007 dla wszystkich leków”).

Dla potrzeb budowy modeli typujących przypadki nadużyć do kontroli wyliczono szereg zmiennych analitycznych. Wykorzystując kreatory dostępne w CLEMENTINE wyprowadzono zmienne, poczynawszy od częstości realizacji recepty przez daną osobę od danego lekarza w konkretnej aptece, aż do zmiennych opartych na wiedzy eksperckiej np. wskaźnik złamania zasad tachyfilaksji (rozdział 3.3.2.3 „Analiza złamania założeń zjawiska tachyfilaksji”).

### **3.3. ZREALIZOWANE ANALIZY — LEKI NARKOTYCZNE (PRZECIWBÓLOWE)**

Nadużycia związane z tą grupą leków wynikają z faktu, iż mogą one stanowić substytut narkotyku lub być źródłem pozyskania morfiny. Leki te wydawane są tylko na recepty i są refundowane przez NFZ. W związku z tym, tanie zakupienie takiego leku w aptece i następnie wprowadzenie go do nielegalnego obrotu może być źródłem potencjalnego nielegalnego dochodu. Ponadto osoby uzależnione mogą próbować podejmować działania mające na celu pozyskiwanie ilości leku większej niż mogłoby to wynikać z ich terapii. W tym obszarze zbudowano modele zarówno typujące podejrzanych pacjentów jak i podejrzanych lekarzy.

#### **3.3.1. Typowanie podejrzanych lekarzy**

##### **3.3.1.1. Model typujący przypadki odstające**

W tym podejściu drążenie danych zastosowano do budowania profili typowych zachowań lekarzy i wyboru tych spośród nich, którzy do utworzonych profili nie pasują. Modelowanie rozpoczyna się od znalezienia charakterystycznych i jednorodnych segmentów lekarzy, a następnie stworzenia ogólnego wzorca dla każdego zidentyfikowanego segmentu.

Na podstawie wzorców następuje selekcja przypadków odstających, które nie pasują do znalezionych reguł. Metoda ta jest podobna do stosowanych na co dzień reguł eksperckich w kontroli, jednak w przypadku drążenia możliwe jest jednoczesne przetwarzanie ogromnej ilości danych na podstawie rozległych charakterystyk zachowań każdego lekarza.

Jeżeli lekarz jest zaangażowany w nielegalne pozyskiwanie refundacji leków narkotycznych, to będzie on dążył do wypisywania recept dla relatywnie dużej liczby powtarzających się pacjentów, którzy mieszkają w różnych powiatach i recepty będą realizowane w różnych aptekach.

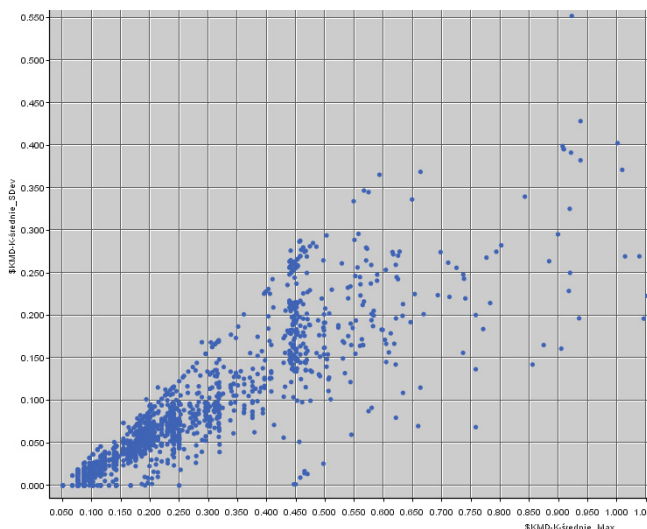
Pacjenci na których wypisywane są recepty mogą nic nie wiedzieć o procedurze, który ma na celu ukrycie nadużycia w ogromie danych. Wykorzystując funkcjonalności CLEMENTINE wykonano następującą analizę:

- wyliczenie wskaźników obrazujących dla danego lekarza:
  - liczbę aptek realizujących recepty w miesiącu,
  - liczbę powiatów w których realizowane są recepty w miesiącu,
  - liczbę kodów terytorialnych w których realizowane są recepty w miesiącu,
  - liczbę recept wydawanych miesięcznie oraz wystąpienie danej grupy ATC leku na receptce;
- w oparciu o wyliczone wskaźniki wykonano analizę skupień – dla jednego skupienia, czyli jego centrum stanowiły przypadki typowe dla realizacji recept w oparciu o wykorzystane dane. Przypadki typowe charakteryzowały się realizacją kilku recept miesięcznie dla 2–4 pacjentów w 1–2 aptekach. Przypadki odstające od tego schematu odpowiadają założonemu, podejrzanemu schematowi nielegalnego działania;
- wygenerowano listę 100 lekarzy uporządkowanych wg miar przyjętych w analizie, czyli według odległości od centrum skupienia. Wygenerowane typy skierowano do kontroli.

Na 18. miejscu wygenerowanej listy znalazł się lekarz już skazany za nielegalny obrót receptami, co szybko dowiodło zasadności koncepcji realizowanej analizy.

### 3.3.1.2. Inne analizy

Przykładem podejrzanego aktywności lekarzy może być również generowanie dużej liczby recept realizowanych głównie w jednej tylko aptece lub wypisywanych głównie na jednego pacjenta.



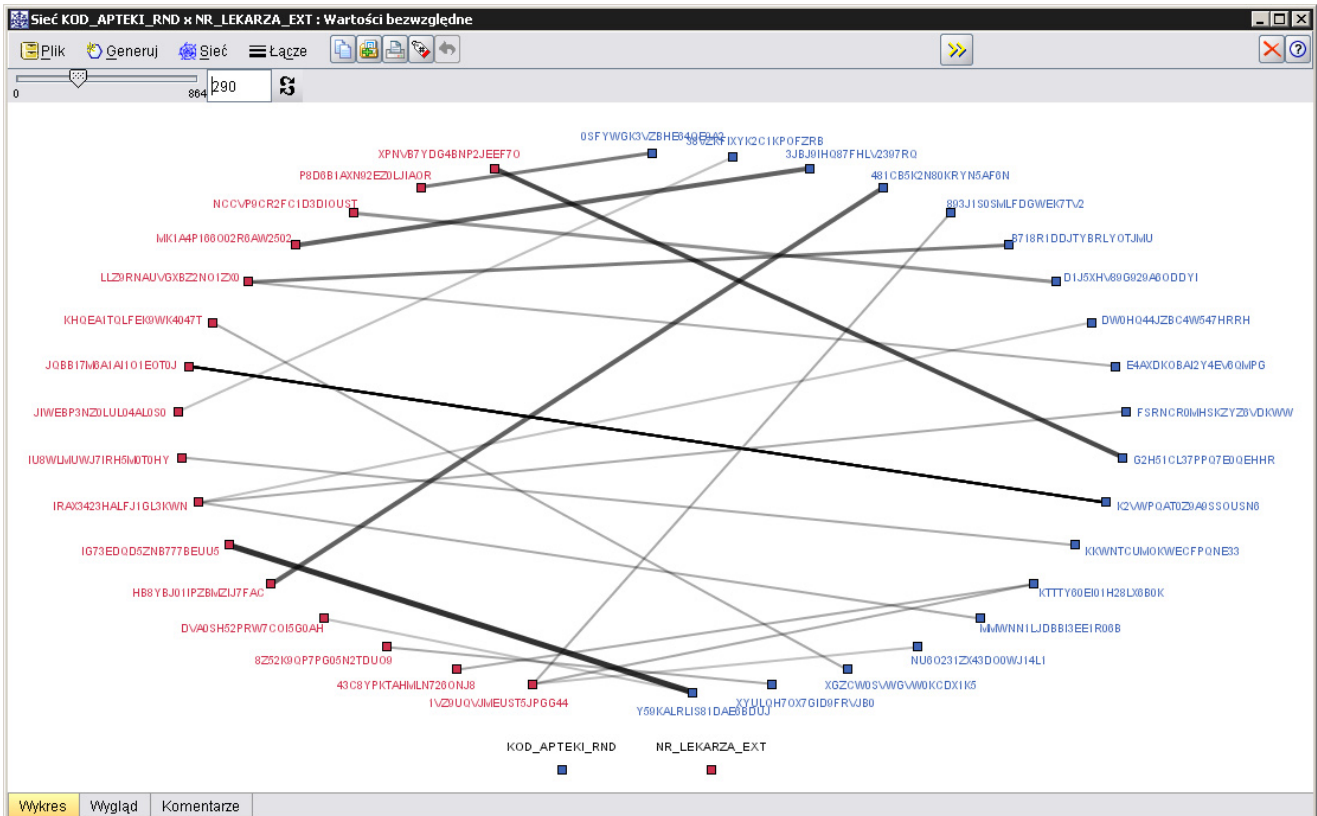
Rysunek 3.

Analiza tych przypadków z wykorzystaniem CLEMENTINE obejmowała:

- agregację danych według identyfikatora lekarza i apteki oraz posortowanie malejąco wyniku;
- agregację danych według identyfikatora lekarza i pacjenta oraz posortowanie malejąco wyniku;
- wizualizację połączeń na wykresach sieciowych.

Odpowiednia wizualizacja zjawisk pozwala na szybkie ustalenie podejrzanego związku między lekarzem a apteką, lekarzem a pacjentem. Wykres na Rysunku 4. obrazuje związki wynikające z liczby recept wypisywanych przez danego lekarza i realizowanych w konkretnej aptece.

Rysunek 4.



### 3.3.2. Typowanie podejrzanych pacjentów

#### 3.3.2.1. Model typujący przypadki odstające

Jeżeli pacjent jest zaangażowany w nielegalne pozyskiwanie leku na użytek własny lub z przeznaczeniem do dalszego obrotu, to będzie uzyskiwał recepty od możliwie różnych lekarzy i realizował je w różnych aptekach, zlokalizowanych w różnych gminach, powiatach.

Dane z identyfikatorem pacjenta są zbierane od 01.07.2007 r. dlatego zbiór zawężono do II półrocza 2007 r. Wykorzystując funkcjonalności CLEMENTINE wykonano następującą analizę:

- wyliczenie wskaźników obrazujących aktywność dla danego pacjenta:
  - liczbę aptek realizujących recepty w miesiącu,
  - liczbę powiatów w których realizowane są recepty w miesiącu,
  - liczbę lekarzy wypisujących recepty miesięcznie,
  - liczbę kodów terytorialnych w których realizowane są recepty w miesiącu,
  - liczbę recept realizowanych miesięcznie oraz wystąpienie danej grupy ATC leku na receptę;
- w oparciu o wyliczone wskaźniki wykonano analizę skupień – dla jednego skupienia, czyli jego centrum stanowiły przypadki typowe dla realizacji recept w oparciu o wykorzystane dane. Typowym schematem okazała się realizacja kilku recept miesięcznie w 1-2 aptekach. Przypadki odstające od tego schematu odpowiadają założonemu, podejrzanemu schematowi nielegalnego działania;

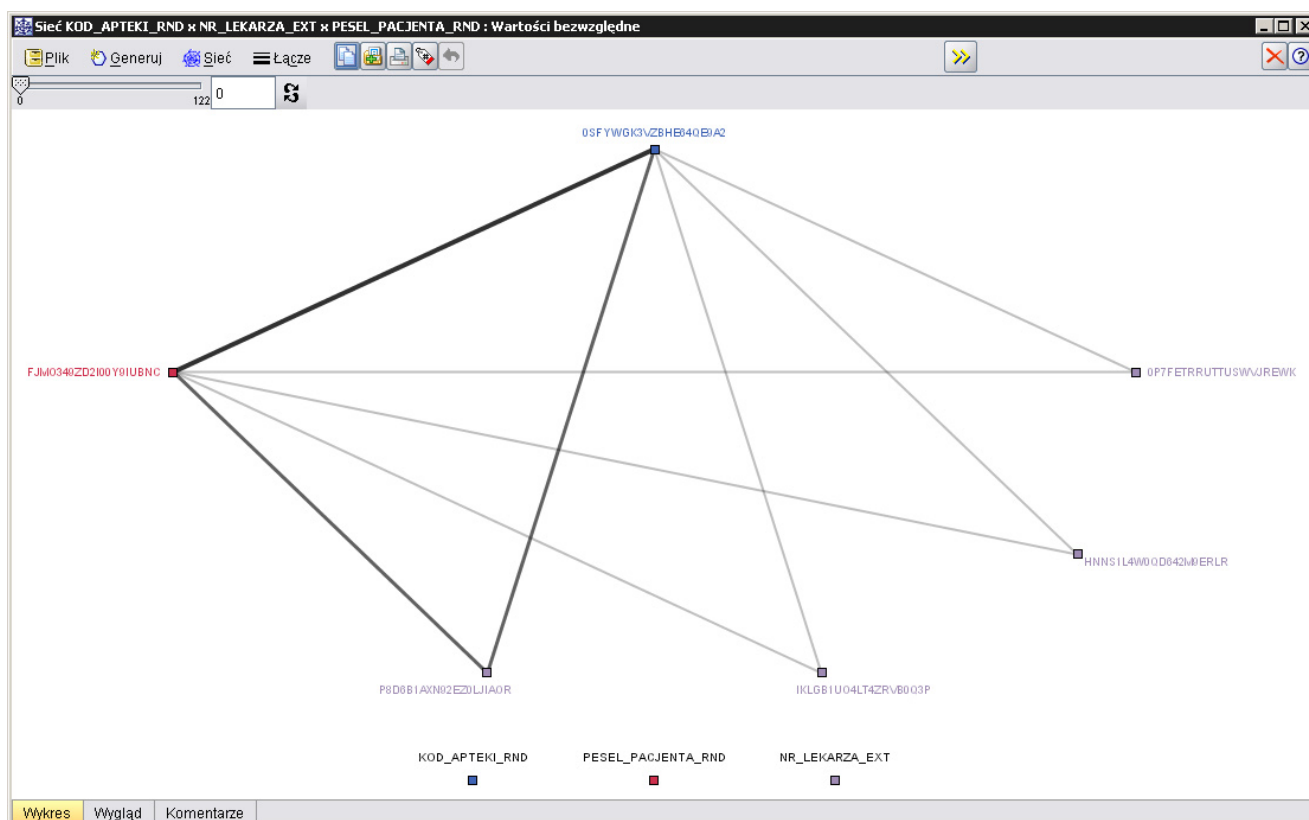
- wygenerowano listę 100 pacjentów uporządkowanych wg miar przyjętych w analizie, czyli według odległości od centrum skupienia. Wygenerowane typy są aktualnie poddawane kontroli.

Jednym z ekstremalnych przypadków okazała się pacjentka, która w ciągu dwóch miesięcy zrealizowała 22 recepty uzyskane od kilku lekarzy, w kilkunastu aptekach znajdujących się w różnych miastach. Jak się później okazało, jej działalność była elementem szerokiego procederu wyłudzenia leków realizowanego przez zorganizowaną grupę na terenie całego kraju.

### 3.3.2.2. Inne analizy

Przykładem podejrzanej aktywności pacjentów jest również duża liczba recept realizowanych głównie w jednej tylko aptece lub wypisywanych głównie przez jednego lekarza. Analiza tych przypadków z wykorzystaniem CLEMENTINE obejmowała:

- agregację danych według identyfikatora pacjenta i apteki oraz posortowanie malejąco wyniku;
- agregację danych według identyfikatora lekarza i pacjenta oraz posortowanie malejąco wyniku;
- wizualizację połączeń na wykresach sieciowych.



Rysunek 5. Odpowiednia wizualizacja zjawisk pozwala na szybką analizę podejrzanych związków między pacjentem a apteką, pacjentem a lekarzem. Wykres na Rysunku 5. obrazuje aktywność wybranego pacjenta w zakresie lekarzy, od których uzyskuje recepty oraz apteki, w której są one realizowane.

### 3.3.2.3. Analiza złamania założeń zjawiska tachyfilaksji

Uzupełnienie analizy pacjentów stanowiła analiza złamania zjawiska tachyfilaksji, które związane jest z procesem przyjmowania w terapii przeciwbólowej coraz większej ilości leku i następnie zmiany leku na silniejszy i ponowny wzrost jego ilości. Jeżeli mamy do czynienia z przypadkami utrzymywania się dużej liczby pozyskiwanego leku i braku zmiany w czasie jego ilości lub spadek tej ilości, może być to przypadek nadużycia.

Wykorzystując funkcjonalności CLEMENTINE obliczono wskaźnik obrazujący zmianę w czasie liczby recept i ilości leku dla pacjenta. Połączenie wyników wszystkich analiz wykonanych ze względu na pacjenta pozwoliło wytypować przypadki pojawiające się na szczycie wszystkich list i przekazać je do kontroli.

	PESEL_PACJENTA	ILOSC_LEKU	LICZBA_RECEPT	ZMN_ILOSC_LEKU	ZMN_LICZBA_RECEPT	LICZBA_TYGODNI
1	2GPSC0NHA8WDO85B1L8T	40.000	34	-0.013	0.000	77
2	RPD23SKE4X9SPNOHARVX	19.000	15	0.000	-0.013	77
3	FS5M4974A76DSK4QUGPB	28.000	26	-0.016	-0.031	64
4	SJYXCD5IU29O2JNAPE2	15.000	8	0.000	0.000	32
5	74YLXS21G0017ABF9PEQ	48.500	32	0.048	0.032	31
6	2T3RDVKI200Y23G7N068	19.000	15	0.016	0.000	31
7	DYDL5CRKK57H428A6X74	16.000	14	0.000	0.000	31
8	Y6YHJWW7EKHKJ2UZ19UO	102.000	52	0.000	0.000	31
9	8IUCN4A57SP4WP829B7	216.500	74	-0.032	-0.032	31
10	RUNASDG1Z1 OWLPOH9S14	261.000	87	-0.194	-0.065	31
11	A2ACIR1ZAR4Q6JVZS862	34.200	35	0.000	0.000	31
12	WEBM51OQG0AA7HTYXOI3	23.000	23	0.032	0.032	31
13	80TTTRZZZ7OHT1P0A97	15.000	25	0.022	0.000	31
14	I4XK4RXOS0LQPUI9OUJR	12.400	55	0.000	-0.032	31
15	XFEKJZVNUI43UPMYL4VS	13.000	7	0.032	0.000	31
16	F80V2E1DRK80CFM4UPQJ	15.000	8	0.000	0.000	31
17	QA86RYMBC3G0W19FY7QW	51.000	27	0.129	0.032	31
18	IJARW6X510TS05UHGU7A	108.000	27	0.000	0.000	31
19	ATIIDGC7ESA86QM5WWM9	13.333	12	0.000	0.000	30
20	3WEKH7NTTGK1XQOWXP...	156.000	38	0.000	0.000	30

Rysunek 6.

W tabeli na Rysunku 6. przedstawiono wyliczone współczynniki związane ze złamaniem założeń tachyfilaksji. W przypadku wartości zmiany mniejszej lub równej zero zachodzi podejrzenie nadużycia.

### 3.4. ZREALIZOWANE ANALIZY — DUBLE RECEPT W ROKU 2007 DLA WSZYSTKICH LEKÓW

Po dokonaniu prostej agregacji okazało się, że identyfikatory części recept pojawiają się wielokrotnie i wielokrotnie odnotowano w bazie refundację z ich tytułu. Wykryte przypadki częściowo mogą być efektem błędów systemowych, zdiagnozowano jednak szereg podejrzanych przypadków związanych np. z realizacją określonych dubli głównie w jednej tylko aptece.

Przypadkiem, który został potwierdzony jako przestępstwo była wielokrotna realizacja tej samej recepty w kilkunastu aptekach przez tego samego pacjenta. Prawdopodobnie oryginalna recepta została skopiowana i następnie realizowana w różnych aptekach i miastach. Dodatkowo okazało się, że ta sama recepta była również realizowana w innych województwach.

Recepta była wypisana na lek przeciwbólowy Tramal, ale nie zaklasyfikowany do leków narkotycznych. Lek ten jednak w połączeniu z niewielką ilością alkoholu daje efekt narkotyku i dlatego jest atrakcyjnym towarem nielegalnego obrotu.



Rysunek 7. Na wykresie (Rysunek 7.) widzimy pacjenta, który zrealizował tę samą receptę w kilku-nastu aptekach.

### 3.5. ZREALIZOWANE ANALIZY — LEKI INSULINOWE W ROKU 2007

Leki insulinowe charakteryzują się wysokimi cenami i stanowią znaczną pozycję w budżecie refundacji leków. Nadużycia w tym obszarze będą zatem związane z dążeniem do uzyskania nielegalnej refundacji. Analizę można wykonać analogicznie do zrealizowanych przy lekach narkotycznych dla potrzeb typowania podejrzanych lekarzy i pacjentów.

W tym wypadku postanowiono przeprowadzić inny typ analizy, tzw. analizę przyspieszenia, czyli poszukiwanie przypadków nagłego wzrostu liczby recept i kwot przekazywanej refundacji. Jednostką analizy są w tym wypadku apteki, które są bezpośrednim beneficjentem refundacji. Wykorzystując funkcjonalności CLEMENTINE wyliczono wskaźniki obrazujące miesięczne zmiany: realizacji liczby recept na leki insulinowe i kwot refundacji.

Na podstawie tej analizy wygenerowano listę aptek, które w ciągu jednego miesiąca odnotowały znaczny wzrost kwot refundowanych. Wytlumaczeniem takiego zjawiska może być aktywność przedstawicieli handlowych firm farmaceutycznych, ale również nielegalnego generowania recept w celu uzyskania refundacji.

Analizę przyspieszenia wykonano również na zbiorze całości refundacji realizowanej w 2007 roku. W wyniku tej analizy odkryto wirtualny punkt apteczny, który istniał tylko w systemie informatycznym i formalnych dokumentach, natomiast nie miał faktycznej siedziby. Punkt okazał się miejscem realizacji nielegalnych recept i wyłudzenia sporych kwot refundacji.

	KOD_APTEKI	X_REFWAG_SUMA	ROK_MIESIAC	ZMN_X_REFWAG_SU...
1	XDC9N030EN8ALOH81 MF0	2596.170	200706	0.000
2	XDC9N030EN8ALOH81 MF0	9854.280	200707	7258.110
3	XDC9N030EN8ALOH81 MF0	11055.050	200708	1200.770
4	XDC9N030EN8ALOH81 MF0	19154.360	200709	8099.310
5	XDC9N030EN8ALOH81 MF0	53815.080	200710	34660.720
6	XDC9N030EN8ALOH81 MF0	14515.940	200711	-39299.140
7	XDC9N030EN8ALOH81 MF0	16299.000	200712	1783.060

Rysunek 8.

Tabela na Rysunku 8. obrazuje nagłe zmiany kwot refundacji w zakresie leków insulino-  
wych dla wybranej apteki.

#### 4. Podsumowanie

Powyżej przedstawiono w telegraficznym skrócie tematykę większości zrealizowanych analiz, należy jednak zwrócić uwagę na fakt, że projekt pilotażowy zrealizowano w ciągu 5 dni i bez przeprowadzenia wcześniejszych przygotowań. Zastosowanie CLEMENTINE pozwoliło na elastyczne wykorzystanie dostępnych danych, przekształcanie ich dla potrzeb analitycznych i ostateczne modelowanie.

Drążenie danych ma szereg zastosowań, a w szczególności w identyfikacji nadużyć w obrocie lekami i ich refundacji. W tym przypadku drążenie danych pomogło ulepszyć tradycyjne strategie kontroli, w celu osiągnięcia lepszych wyników. Przyjazne środowisko analityczne, jakie oferuje CLEMENTINE pozwala bez znajomości zagadnień informatycznych realizować złożone projekty data mining i analizować dane. Efekty projektu udowadniają, że połączenie wiedzy merytorycznej pracowników NFZ i wykorzystanie profesjonalnego oprogramowania może szybko i w relatywnie łatwy sposób przynieść wyraźne odczuwalne efekty.

Grzegorz Zagórny  
Naczelnik Wydziału Gospodarki Lekami  
Śląskiego Oddziału Wojewódzkiego NFZ

Piotr Komornicki  
Prezes  
SPSS POLSKA

*Efekty pilotażu przerosły nasze oczekiwania, jeżeli chodzi o uzyskane rezultaty. Nasi pracownicy oraz konsultanci SPSS Polska wygenerowali szereg list podmiotów do kontroli w zakresie leków przeciwbólowych i insulino-  
wych. Już wstępna weryfikacja pozwoliła potwierdzić kilka przypadków poważnych nadużyć na kwotę przynajmniej kilkuset tysięcy złotych. Oprogramowanie Clementine dowiodło swojej skuteczności w zakresie wykrywania nadużyć i w naszej opinii wykorzystanie go na szeroką skalę z pewnością przyniesie konkretne korzyści dla budżetu a także będzie wpływało prewencyjnie*

”

Grzegorz Zagórny  
Naczelnik Wydziału  
Gospodarki Lekami  
ŚOW NFZ



---

**SPSS POLSKA**  
ul. Ractawicka 58  
30-017 Kraków  
tel./faks 012.636.96.80  
e-mail: [info@spss.pl](mailto:info@spss.pl)  
[www.spss.pl](http://www.spss.pl)  
[www.analizadanych.pl](http://www.analizadanych.pl)  
[www.webmining.pl](http://www.webmining.pl)