



Narodowy Fundusz Zdrowia Wielkopolski Oddział Wojewódzki w Poznaniu

Podsumowanie z przeprowadzenia pilotażu funkcjonowania oprogramowania PASW Modeler (dawniej Clementine) firmy SPSS¹ służącego do poszukiwania przypadków nadużyć w zakresie refundacji leków w WOW NFZ

Zgodnie z zawartym pomiędzy Wielkopolskim Oddziałem Wojewódzkim NFZ, a firmą SPSS POLSKA porozumieniem w dniach 27-29 października 2008 r. przeprowadzono pilotaż oprogramowania analitycznego PASW Modeler. Przed przeprowadzeniem testów wiele czasu poświęcono na zdiagnozowanie problemów analitycznych i sformułowanie potrzeb, przed którymi stoją pracownicy WOW NFZ. Zgodnie z założeniami zainstalowano i uruchomiono oprogramowanie na serwerze Oddziału. Realizacja podstawowych zestawień pozwoliła dokonać oceny pracy programu zarówno pod względem jakościowym jak i wydajnościowym. Pierwsze dwa dni zajęła instalacja PASW Modeler oraz wygenerowanie zbioru analitycznego będącego materiałem wykorzystanym do prac analitycznych. Ze względu na ograniczony czas trwania pilotażu do analizy wybrano jedynie leki z jednej grupy ATC. Spośród zbioru zawierającego około 300 mln rekordów wybrano 55 mln rekordów z lat 2007-2008 co zawężono do 5,5 mln rekordów obejmujących leki z grupy ATC „R” co odpowiada wolumenowi ok. 3 mln recept.

Do pełnej analizy wymagane byłoby przeprowadzenie trwającego do kilku tygodni audytu danych, tak aby móc określić jakość i rzetelność posiadanych danych. W chwili obecnej po wstępnej analizie wątpliwości budzi około 8-10% danych, co jest wartością zbieżną z danymi o wątpliwej jakości, uzyskiwanymi podczas pracy z systemem KS-SIKCH.

W dniu 29 października przeprowadzono agregację danych i przystąpiono do generowania analiz z użyciem oprogramowania PASW Modeler firmy SPSS. Zgodnie z ustaleniami zawartymi w porozumieniu o przeprowadzeniu pilotażu przygotowano zestawienia obrazujące liczby aptek realizujących leki

w analizowanej grupie z informacjami o refundacji na poszczególną aptekę, liczbie pacjentów i lekarzy przypadających na aptekę. Najbardziej interesujące z punktu widzenia analizy było przedstawienie danych w postaci zagregowanej tabeli. Podobnie postąpiono w przypadku analiz dotyczących lekarzy uzyskując dane dotyczące sumy refundacji, liczby recept i pacjentów oraz w ujęciu pacjent-lek-lekarz-aptka. Ciekawe dane uzyskano analizując częstość występowania poszczególnych zdarzeń porównując względne miesięczne wzrosty występowania zdarzenia oraz występowanie pików w stosunku do sąsiadujących przedziałów czasowych. Tego typu obrazowanie zdarzeń nie jest możliwe przy użyciu dotychczas używanego oprogramowania.

Następnie przystąpiono do wykorzystania technik grupowania, wielowymiarowych analiz statystycznych, analiz k-średnich i sieci kohonena w celu stworzenia obrazu „modelowego lekarza” i zbadania listy przypadków odstających, które będzie można poddać kontroli. Analiza skupień i odchyień od centrum pozwala w graficzny i łatwy do interpretacji sposób zobrazować występowanie przypadków odbiegających od średniej np. od modelu średniej ilości refundowanych leków w danym okresie. Poza tym PASW Modeler umożliwia wykorzystanie wielu innych bardziej skomplikowanych narzędzi statystycznych, które po pewnym czasie mogłyby być zastosowane w pracach analitycznych.

PASW Modeler to narzędzie analityczne pozwalające spojrzeć na dane w sposób wielowymiarowy. Pomaga wykrywać wzorce w danych wskazujące na nadużycia finansowe lub niegospodarności zarówno przez pacjentów, lekarzy i apteki, które ze względu na olbrzymie ilości danych w bazach informatycznych OW NFZ, jak i ich rozproszenie, pozostają niezauważalne.

Poszukiwanie nadużyć nieodzwrotnie związane jest z pracą na bardzo dużych zbiorach danych, co wymaga z kolei zastosowania narzędzi sprawdzonych na obszarze projektów data mining. Propozycją

¹ SPSS was acquired by IBM in October 2009

firmy SPSS POLSKA jest zbudowanie Systemu Fraud Detection opartego na oprogramowaniu PASW Modeler oraz procedurach analitycznych umożliwiających generowanie list podmiotów do kontroli. Użytkownicy systemów mogą szybko i sprawnie realizować skomplikowane operacje analityczne na dużych zbiorach danych, bez konieczności posiadania zaawansowanej wiedzy informatycznej z zakresu baz danych, które to następnie mogą być interpretowane przez pracowników Działu Kontroli.

PASW Modeler jest szeroko rozwiniętym narzędziem do poszukiwań, nie proponuje gotowych rozwiązań ale pozwala wychwycić nieprawidłowości „po środku”, a nie tylko w „topowych” aptekach. Jego istotą jest samodzielne budowanie wzorców i analizowanie występujących odchyłeń, przy czym użytkownik musi dokładnie wiedzieć czego poszukuje tj. wcześniej należy określić możliwe mechanizmy powstawania nadużyć.

Umożliwia on-line analizowanie zmian we wcześniej wygenerowanych modelach pod względem wybranych parametrów ze szczególnym uwzględnieniem dynamiki zmian częstości.

Dodatkowej analizie wymagało zbadanie czy realne jest sprawne generowanie (maksymalnie 24 godzin) modelowych zachowań na poziomie lekarz-pacjent-apteka dla całej bazy refundacji i wskazywanie zachowań odbiegających od założeń wg zadanych kryteriów (baza o wielkości wielu setek milionów rekordów). Ograniczenie czasu trwania pilotażu pozwoliło jedynie oszacować czas potrzebny na wygenerowanie pełnych zestawień. Badana grupa leków spowodowała obciążenie serwera w ciągu około 6 godzin, jednak nie była znana realna wydajność oprogramowania na pełnej bazie informacji o refundacji leków. Dysponując licencją na oprogramowanie PASW Modeler w wersji testowej, wydało się celowym dokonanie próbnego wygenerowania na serwerze WOW NFZ zbioru analitycznego obejmującego całą bazę danych dotyczących refundacji leków, co pozwoliło dokładniej oszacować wydajność systemu.

Ostateczne testy wydajnościowe obejmujące całą ściową bazę informacji dotyczących refundacji leków z lat 2007-2009 odbyły się w dniach 25-27 lutego 2009 roku. W tym czasie zdołano wygenerować zestawienie obejmujące ok. 60 mln rekordów grupujących 14 podstawowych cech tabeli źródłowej. Po przeprowadzeniu tej próby można oszacować czas generowania zestawienia na poziomie kilku godzin. Niestety podobnie jak podczas poprzednich testów nie zidentyfikowano ok. 15% danych. Można wstępnie założyć, że analiza wszystkich cech charakteryzujących refundacje leków trwałaby zgodnie z pierwotnym założeniem nie dłużej niż dobę. Nie zdołano natomiast przeprowadzić wszystkich niezbędnych testów celem uzyskania zestawień i wyników, które mogłyby jednoznacznie potwierdzić na tym etapie wzrost skuteczności kontroli realizowanych w oparciu o wyniki uzyskane dzięki PASW Modeler, z wynikami uzyskanymi aktualnie realizowanymi metodami.

Utrudnione jest, z użyciem dotychczas wykorzystywanych narzędzi, prowadzenie dokładnej analizy statystycznej z uwzględnieniem wielu zmiennych dających obraz zachowań związanych z refundacją leków. Ponadto wielkości obsługiwanych baz danych nie pozwalają na w miarę sprawne i nie czasochłonne generowanie skomplikowanych zestawień będących podstawą do późniejszych analiz i wnioskowania, ograniczeniem w tym wypadku jest wydajność serwerów obsługujących bazy recept w WOW NFZ oraz struktury samych baz co stwarzało zresztą istotne problemy podczas pilotażu programu PASW Modeler.

Wszystkie narzędzia informatyczne, którymi posługuje się do tej pory Fundusz, z natury służą do generowania zestawień i raportów zawierających dane o różnym poziomie szczegółowości.

W pewnym momencie pojawiają się jednak problemy ze sprawną i poprawną analizą uzyskanych danych pod względem wychwytywania zjawisk nietypowych zachodzących w refundacji leków.

Rozwiązaniem wymienionych problemów może być testowane w WOW NFZ, jak i wcześniej w Śląskim Oddziale Wojewódzkim NFZ, narzędzie analityczne PASW Modeler. Prezentowane przez program PASW Modeler możliwości spełniają funkcjonalności potrzebne w WOW NFZ, jednak koniecznym wydaje się być przeprowadzenie szczegółowego audytu danych tak, aby ograniczyć lub wyeliminować informacje, które nie są obecnie identyfikowane.



Naczelnik
Wydziału Gospodarki Lekami
Barbara Mroczyk