WYKRYWANIE NADUŻYĆ I PRANIA BRUDNYCH PIENIĘDZY

Ton Kuijlen
konsultant StatSoft Polska Sp. z o.o.

Grzegorz Migut
StatSoft Polska Sp. z o.o.

Wstęp

Każdego dnia w systemie bankowym rejestrowane są tysiące różnego rodzaju transakcji. Znakomita większość z nich jest zapisem typowych operacji związanych z legalnie podejmowanymi działaniami. Niestety pewna ich część reprezentuje działania patologiczne – pranie brudnych pieniędzy, oszustwa bankomatowe itp. Dla instytucji pragnących cieszyć się dobrą reputacją i wiarygodnością, umiejętność wychwytywania tego typu transakcji oraz zapobieganie im jest koniecznością. Potrzeba podejmowania tego typu działań wynika również z uwarunkowań prawnych. Wraz z nowelizacją ustawy o przeciwdziałaniu praniu pieniędzy, na instytucjach bankowych spoczął obowiązek analizy rejestrowanych przez nie transakcji pod kątem ich legalności. Nie bez znaczenie są również zalecenia Komitetu Bazylejskiego.

Monitorowanie transakcji w celu wykrywania nadużyć i prania pieniędzy jest jednym z obszarów szerokiego wykorzystania technik zgłębiania danych (data mining). Dzięki systemom wykorzystującym zaawansowane metody analizy danych możliwe jest nie tylko wykrywanie transakcji zarejestrowanych przez system, ale również (co nie jest już wymogiem ustawowym) stosowanie reguł w czasie rzeczywistym dla transakcji, które są w toku (np. wypłata z bankomatu).

Nowoczesne podejście do budowy systemów służących do wykrywania nadużyć i prania pieniędzy polega na zdobywaniu wiedzy zarówno na temat transakcji, jak i osób je wykonujących oraz pracowników banku. Wykorzystanie technik analitycznych umożliwia w dużym stopniu zautomatyzowanie operacji wyszukiwania nielegalnych transakcji. Dzięki temu podejściu również organizacje posiadające olbrzymią ilość klientów (ponad 2 miliony) są w stanie wykryć podejrzane transakcje, klientów, czy pracowników, co wcześniej było praktycznie niemożliwe ze względu na olbrzymią ilość danych. Podejście to umożliwia wykonanie podziału klientów ze względu na poziom ryzyka wykonania przez nich nielegalnej operacji. Poprzez identyfikację grup ryzyka (np. grupy niskiego, średniego i wysokiego ryzyka) wykrywanie nadużyć i innych nielegalnych operacji może skupić się przede wszystkim na najbardziej zagrożonych osobach. Oczywiście by system służący do

wykrywania prania pieniędzy działał sprawnie, konieczny jest także udział odpowiednio przygotowanej grupy analityków, do której należy ostateczna decyzja odnośnie wykrytych nieprawidłowości.

Działanie systemu służącego do wykrywania nadużyć i prania brudnych pieniędzy opiera się na formułowaniu i stosowaniu reguł określających, czy dana transakcja jest legalna czy nie. Reguły te mogą się opierać na tak zwanych mocnych kryteriach, np. każda operacja pieniężna powyżej 15 000 € musi zostać sprawdzona. Inna grupa reguł opiera się na śledzeniu anomalii – wyłapywane są te transakcje, które znacząco odbiegają od standardowego lub średniego zachowania klientów. W niektórych rzadkich sytuacjach stosuje się „ręczne" sprawdzanie, dzieje się tak w przypadkach, które cechują się wysokim ryzykiem, lecz trudno je wychwycić poprzez standardowe profilowanie.

Monitorowanie klientów oraz zawieranych przez nich transakcji wymaga, by wzorce opisujące ich zachowanie były budowane na podstawie danych z długiego okresu czasu (przynajmniej roku). Skuteczne stosowanie zbudowanych reguł wymaga podejścia zakładającego ich nieustanne dostosowywanie do zmieniających się warunków. W wielu przypadkach dana reguła wymaga aktualizacji już po bardzo krótkim czasie jej obowiązywania.

Metody prania pieniędzy

Działalność przestępcza polegająca na praniu pieniędzy opiera się na jednym z dwóch rodzajów aktywności. Pierwszy rodzaj polega na przemieszczaniu środków pomiędzy różnymi lokalizacjami, drugi polega na zmianie ich formy z gotówki na przykład na czeki podróżne lub wartościowe przedmioty. Celem przestępców jest wprowadzenie brudnych pieniędzy do oficjalnego obiegu za pomocą działań niewyróżniających się na tle zachowań legalnie działających osób.

W procederze prania pieniędzy możemy wymienić trzy zwykle występujące ze sobą fazy działania, których łączne występowanie ma na celu zmniejszenie ryzyka wykrycia rzeczywistego źródła pochodzenia brudnych pieniędzy. Oczywiście każda z osobna nosi znamiona przestępstwa, tak więc nawet wystąpienie tylko jednej z wymienionych poniżej faz, nie powoduje zmiany statusu nielegalnego przedsięwzięcia [2].

Umiejscawianie (*Placement*) polega na wprowadzeniu nielegalnie zdobytych środków do systemu bankowego lub legalnego przedsięwzięcia. Fazę tę często poprzedza przemyt brudnej gotówki, co ma na celu utrudnienie ewentualnego zlokalizowania jej prawdziwego źródła. Znamienną cechą tej fazy jest fakt, że nielegalnie zdobyte środki mają najczęściej formę gotówki. Podejmowane działania cechują się dużą prostotą i zwykle są krótkoterminowe.

Maskowanie (*Layering*) - drugie stadium procederu - polega na oddzieleniu pieniędzy od ich nielegalnych źródeł poprzez wykonywanie szeregu finansowych transakcji, takich jak na przykład transfer pomiędzy kontami w banku, wymiana na czeki podróżne itp. Poprzez tego rodzaju działania sprawca ma nadzieję uniemożliwić lub przynajmniej utrudnić wykrycie powiązania środków z prawdziwym źródłem ich pochodzenia.

Integracja (*Integration*) polega na łączeniu środków z legalnie zdobytymi pieniędzmi lub dostarczaniu budzących zaufanie wyjaśnień co do ich pochodzenia. Wśród narzędzi używanych do osiągania integracji wykorzystuje się fikcyjne sprzedaże nieruchomości, fikcyjne podmioty gospodarcze, współpracę banków zagranicznych itp.

Oto przykłady podejrzanych transakcji zarejestrowanych w systemie bankowym.

- Klient kupuje drogi produkt (>15 000 €) i na przykład wstępnie płaci 10 000 €, a następnie wraca tego samego dnia, by zapłacić pozostałą kwotę również w formie gotówki.
- Osoba pochodząca na przykład z jednego z państw rozwijających się, posiadająca prywatne konto, przelewa dużą sumę pieniędzy ze swoich kont znajdujących się za granicą, a następnie wybiera całą gotówkę w ciągu jednego dnia, dokonując wielokrotnych wypłat przy pomocy bankomatu.
- Osoba dokonuje transakcji bardzo dla siebie niekorzystnej.
- Klient dokonuje transakcji na kwotę tuż poniżej kwoty, która automatycznie jest raportowana w systemie.

By lepiej sterować procesem prania brudnych pieniędzy, wynajdywane są wciąż nowe metody mające na celu zakamuflowanie źródeł pochodzenia nielegalnie zdobytych pieniędzy. Różnorodne techniki prania pieniędzy występować mogą w każdej z opisanych faz. Ich katalog jest bardzo obszerny, poniżej prezentowane są wybrane z nich [2].

Mieszanie. Jedną z najpopularniejszych technik prania pieniędzy jest łączenie brudnych pieniędzy z dochodami legalnego podmiotu gospodarczego, zwane „mieszaniem". Przedsiębiorstwo prowadzi działalność gospodarczą i nastawione jest na osiąganie zysków; dlatego dołączanie do ich dochodów brudnych pieniędzy, np. na ich rachunki bankowe, nie wzbudza podejrzeń, jakie stwarzać może np. zakup drogiego samochodu dokonany przez bezrobotnego.

Smurfing. Często wykorzystywaną techniką jest tzw. *smurfing* polegający na wykorzystywaniu dużej liczby drobnych pośredników (nazywanych *smurfami*), którzy nabywają czeki bankowe i inne papiery wartościowe na okaziciela na wartość poniżej granicy obligującej do identyfikacji albo też dokonują licznych i wielokrotnych transakcji w wielu różnych instytucjach. *Smurfy* wykorzystuje się także do zamiany banknotów pochodzących z przestępczego procederu, które zwykle są w niskich nominałach, na wysokie nominały. Wykorzystywać można w tej technice np. kantory czy kasyna gier, gdzie gotówka jest wymieniana na żetony, a następnie na banknoty o najwyższych nominałach lub czeki.

Kredyt dla siebie. Mechanizm tego procederu polega na zaciągnięciu dwóch kredytów. Jednego legalnego z banku, drugiego fikcyjnego od powiązanej (najlepiej zagranicznej) spółki. Kredyt legalny spłacany jest wraz odsetkami z nielegalnych środków pochodzących oficjalnie z drugiego kredytu. Dzięki temu przestępcy zyskują legalnie zdobyte środki i zadłużenie wobec samych siebie (jako fikcyjnej firmy).

Transferpricing. Do prania pieniędzy wykorzystuje się transakcje w handlu międzynarodowym. Najpowszechniejsza technika określana jest jako *transferpricing*. Wykorzystuje

się w niej powiązane ze sobą podmioty gospodarcze, które dokonują nadfakturowywania lub niedofakturowywania eksportu czy importu. Dla przykładu transferu nielegalnie zdobytych środków można dokonać, zawierając kontrakt na import towarów z zagranicy ze znacznie zawyżoną ceną.

Transakcje puste – tego typu transakcje związane są z fałszowaniem faktur dotyczących całkowicie fikcyjnych operacji handlowych.

Budowa systemu do wykrywania nadużyć

System tego typu został z powodzeniem wdrożony w jednym z banków holenderskich. W okresie poprzedzającym budowę systemu bank stosował pewne zdefiniowane reguły, które były wykorzystywane w celu wykrywania nielegalnych operacji przez specjalnie przeszkolonych pracowników. Mechanizm stosowania reguł uruchamiany był nocą dla wszystkich transakcji zarejestrowanych w ciągu doby. W wyniku przyjętej procedury podejrzane transakcje wykrywane były dopiero 24 godziny po ich zarejestrowaniu w systemie.

W związku z regulacjami rządowymi oraz rosnącą liczbą nielegalnych transakcji bank zdecydował się na wdrożenie nowego systemu, który mógłby umożliwić wcześniejsze wykrycie tego typu transakcji. Oczekiwano, że przy pomocy nowego systemu korzystającego z zaawansowanych technik data mining możliwe będzie bardziej efektywne działanie. Modele zawarte w planowanym modelu miały odzwierciedlać wiedzę specjalistów zajmujących się wykrywaniem nadużyć i przypadków prania brudnych pieniędzy. Kluczowa była też możliwość natychmiastowego ich stosowania.

By pomyślnie zrealizować obraną strategię działania, konieczne było zorganizowanie danych w spójny sposób w odniesieniu do relacji pomiędzy osobami i organizacjami będącymi przedmiotem analizy. Dlatego też pierwszym etapem budowy systemu było opracowanie i wdrożenie bazy danych, do której zapisywać miano transakcje, dla których istniało podejrzenie, że są nielegalne. Baza ta działała niezależnie od systemu transakcyjnego i zbudowana została wyłącznie do celów analitycznych. Zaraz po jej wdrożeniu została załadowana wszystkimi wykrytymi dotychczas przypadkami nielegalnych transakcji, działań oraz osób i organizacji, które ich dokonywały.

Nowe transakcje identyfikowane były poprzez stosowanie dla nich reguł zawartych w tak zwanej bazie reguł. Jeśli transakcja została określona jako podejrzana, na stanowisku specjalisty zajmującego się nielegalnymi transakcjami otwierało się okno alarmowe. Opierając się na wyświetlonej regule decydującej o włączeniu alarmu oraz informacji na temat transakcji, osobie, która ją wykonała, oraz odbiorcy, specjalista decydował, czy transakcja ta powinna zostać sprawdzona. Jeśli tak, transakcja ta zapisywana była do bazy w celu przeprowadzenia dodatkowych analiz. Na podstawie zdarzeń będących fałszywymi alarmami oraz przypadkami prawidłowej klasyfikacji nielegalnych działań, system stopniowo uczył się, doskonaląc swoje możliwości klasyfikacyjne.

Głównymi elementami zbudowanego systemu były:

- Analityczna baza danych zawierająca wszystkie przypadki nadużyć oraz osoby podejrzane i wymagające dodatkowych ustaleń.
- Mechanizm używany do stosowania reguł służących do wykrywania podejrzanych przypadków (mogą być to reguły określone przez rząd, ekspertów oraz określone przy pomocy analizy data mining). Reguły były uaktualniane regularnie w związku ze zmianami wzorców nielegalnych zachowań.
- System alarmów uruchamianych na komputerze eksperta, związanych z wystąpieniem podejrzanej transakcji. Po wyświetleniu tego typu alarmu ekspert decydował, czy należy daną transakcję wyjaśniać. Jeśli tak, dane dotyczące transakcji były kopiowane do bazy danych zawierającej przypadki nadużyć i prania brudnych pieniędzy.
- System do analizy danych i data mining, przy pomocy którego baza danych była regularnie analizowana w celu wykrycia nowych wzorców lub reguł.

Wdrożony system opierał się na systemach *STATISTICA Data Miner* oraz *SEWSS*. Narzędzia analityczne zawarte w *STATISTICA Data Miner* umożliwiły opracowanie reguł stosowanych do wykrywania przypadków nielegalnej działalności. Dzięki monitorom *SEWSS* możliwe było bieżące analizowanie rejestrowanych transakcji oraz alarmowanie w sytuacji wystąpienia anomalii.

Wykorzystanie metod data mining do wykrywania nielegalnych operacji

Aby ustalić mechanizm prania pieniędzy, nie wystarczy przeanalizowanie jednej transakcji. Zwykle proceder składa się z kilku powiązanych ze sobą operacji, których wzajemna zależność nie jest na pierwszy rzut oka oczywista. Umiejętność odnalezienia tych powiązań może być więc kluczowa dla wykrycia mechanizmu nielegalnego procederu. Poniżej przedstawiono dwa przykłady wykorzystania metod data mining, których użyto podczas budowy systemu w celu ustalenia związków pomiędzy podejrzanymi transakcjami oraz osobami, które je wykonały.

Analiza połączeń w wykrywaniu prania pieniędzy

Dane zawarte w stworzonej bazie danych poddano analizie za pomocą analizy połączeń i asocjacji. Ta interakcyjna technika służąca do określana i testowania graficznej sieci połączeń została zastosowana jako pierwszy krok w budowie systemu data mining. Wykorzystana metoda umożliwia odnajdywanie związków, relacji oraz krytycznych połączeń i powiązań pomiędzy podejrzanymi osobami czy transakcjami. Analiza pozwala także na identyfikację wzorców powiązań oraz pojawiających się nowych grup osób biorących udział w nielegalnym procederze. W wyniku działania metody otrzymywany jest wykres w postaci siatki zawierającej węzły symbolizujące osoby i organizacje oraz połączenia pomiędzy węzłami, które symbolizują powiązania bądź transakcje. Jeśli pomiędzy danymi dwoma obiektami istnieje powiązanie, wtedy na wykresie będą one

połączone linią. Im silniejsza jest relacja, tym grubsza jest linia i odwrotnie. Najsłabsze relacje są przedstawiane w formie linii przerywanej. Dzięki wizualizacji poszczególnych obiektów oraz połączeń między nimi możliwe było ustalenie i zrozumienie siły tych relacji oraz częstości poszczególnych kontaktów oraz nowych ukrytych powiązań.

Przedmiotem analizy może być na przykład zbiór transakcji gotówkowych pomiędzy określonym lokalnym i zagranicznym kontem bankowym, zbiór transakcji z innego państwa i inne powiązane komercyjne i prywatne interakcje. Wydarzeniami mogą być też wpłaty lub wypłaty z konta.

Wstępna analiza danych dla analizy połączeń

Zanim możliwe było skonstruowanie sieci połączeń i istotnych powiązań, konieczne było kompletne zrozumienie natury danych, które były podstawą analizy. Podobnie jak we wszystkich zadaniach data mining odpowiednie przygotowanie danych do analizy jest jednym z głównych zadań, mających kluczowy wpływ na uzyskane wyniki. Również w tym przypadku transakcyjna baza danych, z której przenoszono dane do bazy analitycznej, zawierała niekompletne i niespójne informacje oraz wielokrotne wystąpienia tych samych rekordów. Tego typu błędy wynikały z faktu, że nie była ona budowana do celów analitycznych, lecz by maksymalnie przyspieszyć przetwarzanie transakcji. Baza danych zawierająca dane transakcyjne związane z obsługą konta zawierała na przykład różne imiona lub numery kont dla tych samych osób (w rzeczywistości posiadających jedno konto) lub różne osoby pod jednym identyfikatorem. By móc prawidłowo określić powiązania pomiędzy poszczególnymi osobami, konieczna była najpierw poprawna identyfikacja wszystkich osób. Proces wyjaśniania niejasności i konsolidacji danych był bardzo czasochłonny, jednak konieczny przed podjęciem jakichkolwiek działań odnośnie zasadniczej części analizy.

Określenie odpowiedniego poziomu szczegółowości danych było kolejną bardzo ważną częścią przygotowania danych do analizy połączeń. Na przykład w bazie danych zawierającej transakcje związane z operacjami finansowymi pożądany poziom szczegółowości może być określony przez osobę fizyczną, grupę, dział czy firmę. Kluczowym czynnikiem była w tym wypadku wiedza pracowników banku zajmujących się wykrywaniem nadużyć, którą wykorzystano w celu określenia najbardziej odpowiedniego pod kątem analizy poziomu szczegółowości danych.

Reprezentacja i konfiguracja osób i organizacji, które dokonują podejrzanych transakcji i są opisane za pomocą różnych informacji zawartych w danych, polega na dwóch operacjach: konsolidacji i wyjaśnianiu niejasności. Dokonano na przykład konsolidacji wielokrotnych transakcji w celu oceny aktywności poszczególnych obiektów (osób, organizacji itp.). Z drugiej strony przeprowadzono operację łączenia i oczyszczania danych w celu poprawy błędów występujących w danych. Przez pewien czas zajmowano się profilowaniem obiektów oraz odkrywaniem anomalii mogących oznaczać oszustwo lub inną nielegalna aktywność. Wymagało to często łączenia rekordów w celu odkrywania wzorców pewnych unikalnych zachowań. Głównym celem było takie sformatowanie danych, by była możliwość identyfikacji istotnych obiektów z transakcyjnej bazy danych.

Na przykład wpłaty na konto bankowe zostały zorganizowane pod kątem możliwości sprawdzenia tej aktywności odnośnie prania brudnych pieniędzy.

Przeprowadzone badania wykazały, że powinno się podzielić podejrzane transakcje na trzy podgrupy:

- operacje pomiędzy firmami (B2B),
- operacje pomiędzy firmami a osobami fizycznymi (B2C),
- operacje pomiędzy osobami prywatnymi (C2C).

Po zakończeniu przygotowania danych do analizy wykonano szereg analiz połączeń.

Analiza połączeń w praktyce

Analiza połączeń została wykorzystana do wykrycia powiązań pomiędzy rekordami w bazie danych zawierającej olbrzymią liczbę obiektów reprezentujących nielegalne transakcje i związane z tym podmioty i osoby. Działania były ukierunkowane na identyfikację nielegalnych zachowań oraz pranie brudnych pieniędzy dzięki określeniu istniejących powiązań pomiędzy podejrzanymi transakcjami i osobami.

Analizę połączeń rozpoczęto od identyfikacji numerów kont, dla których podejrzewano występowanie nielegalnego transferu pieniędzy (z konta lub na konto). Dala tych obiektów ustalano siatkę powiązanych z nim kont i ich właścicieli. System wyłapywał nie tylko osoby dokonujące nielegalnych transakcji, ale również osoby z nimi powiązane. W ten sposób możliwe było identyfikowanie powiązań pomiędzy poszczególnymi kontami i proste ustalenie kręgu podejrzanych osób.

W niektórych przypadkach podejrzane osoby mogą nie używać kont, przy pomocy których dokonywały w przeszłości nielegalnych operacji, ale mogą wykonywać przelewy na konto używane przez inną powiązaną z nią osobę. Tego typu informacja może zaalarmować specjalistę bankowego o możliwości nielegalnej działalności. Używając analizy połączeń, możemy tworzyć sieci zawierające informacje o tym, kto kontaktował się z kim, i przedstawić te powiązania na cyfrowej mapie, dzięki której o wiele łatwiej rozpoznać wzorce zachowań, które mogłyby zostać niezauważone za pomocą tradycyjnej analizy.

Analizę połączeń wykorzystano również w celu wykrycia potencjalnie podejrzanych kont analizując wykonywane poprzez nie transakcje w połączeniu z informacjami o używanej przez właściciela karcie kredytowej. Analizując konta powiązane z oszustwami dokonanymi za pomocą karty kredytowej, wskazywano, którzy z nowych klientów mogą dokonywać oszustw za jej pomocą. W tym celu wybrano konta, które były powiązane z kartami kredytowymi używanymi w nielegalny sposób, a następnie szukano nowych kont zawierających te same informacje dotyczące karty kredytowej.

Analiza połączeń była także wykorzystana do wykrywania nadużyć przez wskazanie podejrzanych zmian zarejestrowanych zaraz po założeniu konta. Podejrzany może być na przykład użytkownik konta, który zmienił adres w ciągu pierwszego tygodnia jego użytkowania. Analiza została również wykonana dla różnego rodzaju danych podawanych przez właścicieli kont – na przykład domowy numer telefonu, numer identyfikacji

podatkowej itp., w celu znalezienia połączeń pomiędzy nimi a kontami zidentyfikowanymi jako podejrzane.

Podsumowanie

Dzięki gruntownemu przygotowaniu danych i wstępnej ich analizie następująca po niej analiza połączeń i asocjacji umożliwiła identyfikację sieci podejrzanych transakcji z zakresu nadużyć i prania brudnych pieniędzy. Opierając się na tych informacjach, można było sformułowć reguły umożliwiające monitorowanie transakcji w celu wykrywania podejrzanych działań.

Grupowanie przypadków prania pieniędzy za pomocą SOM

Zadaniem tego studium było sprawdzenie, czy SOM (sieci neuronowe Kohonena) mogą zostać użyte do wykrycia powiązań i połączenia ze sobą przypadków prania brudnych pieniędzy w oparciu o opisy tego typu przypadków. Rezultatem analizy była mapa topologiczna (siatka), w której każda komórka reprezentuje jedno skupienie opisujące przypadek lub grupę podobnych przypadków prania brudnych pieniędzy. Idealnym rozwiązaniem byłaby mapa, w której każda komórka reprezentuje typ powiązany z pojedynczym przypadkiem przestępstwa. Sąsiadujące komórki mapy powinny zawierać opisy innych przypadków prania brudnych pieniędzy wykazujące z nimi podobieństwo. Przed analizą konieczne było zrobienie rozróżnienia pomiędzy osobami fizycznymi a firmami ze względu na różnice w danych opisujących nadużycie. Poniżej opisano studium przeprowadzone dla osób fizycznych.

Przed zasadniczą częścią analizy przeprowadzono działania polegające na wyborze zmiennych do analizy, kodowaniu i oczyszczaniu zmiennych. Interpretacja znaczenia uzyskanej na podstawie analizy mapy topologicznej oraz jej wartość merytoryczna została niezależnie oceniona przez specjalistów nienależących do grupy zajmującej się opracowywaniem modeli.

Wybór zmiennych, czyszczenie danych oraz kodowanie

Podstawą analizy były dane zawarte w analitycznej bazie danych zawierającej przypadki transakcji opisujących pranie pieniędzy. Zapisywane w bazie rekordy zawierały określone informacje personalne, takie jak: nazwisko, adres, ID oraz inne dane administracyjne, dodatkowo występowało tam szereg pól tekstowych służących do opisu rodzaju nielegalnej operacji czy metody działania przestępców.

Pierwszym zadaniem, którym zajęto się z wielką ostrożnością i uwagą, było przekodowanie opisowych zmiennych na zmienne liczbowe wymagane przez sieć SOM.

Kolejny problem, na jaki napotkano, zajmując się zmiennymi, wynikał z różnic zakresów wartości poszczególnych zmiennych ciągłych. Sieci SOM są metodą wrażliwą na różnice w zakresach zmiennych. Jeśli zmiennych tych nie sprowadzilibyśmy do jednego, wspólnego zakresu wartości, wtedy zmienna przyjmująca większe wartości miałaby większy wpływ na ustalanie parametrów modelu, niż zmienna przyjmująca niewielkie

wartości, mimo iż rzeczywisty wpływ obu zmiennych na wyjaśnianie danego zjawiska mógłby być podobny. By zminimalizować ewentualną utratę części informacji zawartych w zmiennych, zdecydowano się na wykorzystanie kodowania binarnego. Zmienne te zostały rozbite na pewną ilość przedziałów. Każdemu przedziałowi odpowiadała jedna zmienna binarna. Zmienna odpowiadająca temu przedziałowi, w którym mieściła się dana wartość, przybierała wartość 1, zmienne odpowiadające pozostałym przedziałom przyjmowały wartość 0.

Podczas rozważanego trzyletniego okresu zanotowano 1 100 przypadków prania pieniędzy. Każdy opis był reprezentowany przez grupę atrybutów, takich jak: wiek, kwota, ilość osób zamieszanych, powiązania z innymi podejrzanymi, numer segmentu, do którego należał rodzaj transakcji itp. Po zakodowaniu zmiennych na reprezentację binarną dało to 46 zmiennych.

Budowa modelu

Podczas ustalania parametrów sieci (uczenia) stosowano nieuporządkowany zbiór przypadków uczących (w naszym przypadku było to 1100 opisów prania pieniędzy), z których każdy zwierał 46 zmiennych. Podczas uczenia iteracyjnie prezentowano sieci wszystkie przypadki, które były przez nią grupowane w określone skupienia.

SOM może być traktowana jako metoda redukująca wielowymiarową przestrzeń wektorową (wielowymiarowy problem) do dwóch wymiarów, w tym przypadku 46 do 2. W rezultacie redukcji do dwóch wymiarów uzyskuje się raczej topologiczną niż przestrzenną mapę. Na mapie tej przypadki zaklasyfikowane jako takie same trafiają do wspólnego skupienia, natomiast przypadki podobne są rozmieszczane w skupieniach sąsiednich. Zdecydowano się na skonstruowanie mapy o wymiarach 9 na 12. W każdej komórce algorytm mógł umieścić dowolną liczbę przypadków, w zależności tego, jak podobne były do siebie.

Wyniki działania modelu

Nie wszystkie komórki na mapie zostały wypełnione - 9 z nich było pustych W celu ułatwienia interpretacji zbudowanej mapy zastosowano proces nadawania symbolicznej nazwy każdemu ze zbiorów. Określano średnią dla każdego z atrybutów dla danego skupienia. Jeśli średnia wartość była większa niż 0.5, wtedy nazwę tego atrybutu umieszczano w nazwie skupienia.

Uzyskana w ten sposób mapa razem z poszczególnymi przypadkami przypisanymi do mapy została przekazana grupie specjalistów, którzy nie brali udziału w procesie jej budowy w celu niezależnej weryfikacji. Mieli oni dostęp do dodatkowych informacji, z których nie korzystano podczas budowania mapy. Poszczególne skupienia były analizowane indywidualnie bez dostępu do połączeń pomiędzy nimi i sąsiadującymi klastrami.

Z ocenianych skupień jedno zawierało niewystarczającą ilość informacji, by wydać o nim sąd. Dla pięciu nie wykryto oczywistych związków pomiędzy przypadkami prania

pieniędzy, jakie zawierały. Natomiast pozostałe skupienia zawierały rzeczywiście ściśle powiązane ze sobą przypadki nadużyć.

Byliśmy w stanie poprawić rezultat analizy przez połączenie ze sobą niektórych sąsiadujących skupień, ponieważ skonstruowana mapa 9-12 wydawała się nieco za duża. Dlatego też połączono skupienia przy pomocy hierarchicznej analizy skupień.

Dodatkowym zastosowaniem zbudowanego modelu było wykorzystanie go w sytuacjach, gdy wiedziano, że dwie nielegalne transakcje są ze sobą powiązane, natomiast trafiły one do dwóch różnych skupień, na przykład (0,0) i (4,4). Jest bardzo możliwe, że uważna analiza przypadków znajdujących się w skupieniach pomiędzy nimi może wyjawić inne transakcje powiązane z tymi dwoma.

Literatura

1. Gontarz A., *Bezpiecznie na bank*, Computerworld 8 czerwca 2004.
2. Ministerstwo Finansów, *Zadania instytucji obowiązanych w przeciwdziałaniu praniu pieniędzy,* Generalny Inspektor Informacji Finansowej, Warszawa 2001.
3. Westphal Ch., Blaxton T., *Data Mining Solutions, Methods and Tools for Solving Real-World Problems*, John Wiley & Sons, Inc, New York 1998.