

Test: Veeam
Availability Suite v8 s. 68

ADFS 3.0 – wdrażanie
usług federacyjnych s. 22

6

Miesięcznik informatyków i menedżerów IT.

www.it-professional.pl

2015

IT professional

Nr 6 (43) czerwiec 2015

Cena 33,00 zł (w tym 5% VAT)

WIRTUALNY STORAGE

s. 10

► Korzyści z wdrożenia rozwiązań Software-Defined Storage. Implementacja założeń SDS proponowana przez VMware w nowej wersji VSAN. Innowacje w magazynowaniu danych w Windows Server 2012 R2.



s. 59

Opis przedmiotu zamówienia na software

Prawidłowe przygotowanie opisu.
Najczęściej popełniane błędy

s. 51

Mistrzowie danych

Kim są mistrzowie danych?
Wymagane kompetencje, polskie realia

s. 40

Linux pod kontrolą AD

Zarządzanie stacjami roboczymi
za pomocą GPO i Centrify Server Suite

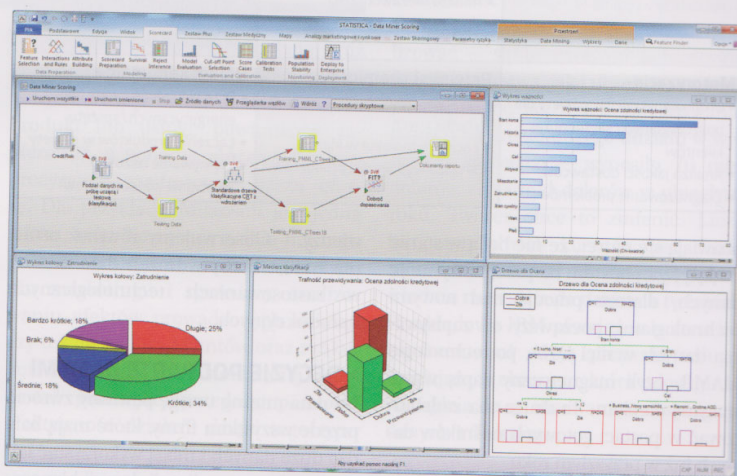
DATA MINING – ZASTOSOWANIA

Eksploracja danych to wykorzystywanie komputerów do znajdowania ukrytych dla człowieka prawidłowości w wielkich zbiorach danych. Wyjaśniamy, dlaczego warto zainteresować się data miningiem.

Grzegorz Kubera

Eksploracja danych (data mining) to automatyczne przesiewanie dużych ilości danych w celu uzyskania cennych informacji i formułowania prognoz. Podobne procesy były stosowane przez statystyków od dziesiątków lat, choć oczywiście nosiły inną nazwę. Termin data mining pojawił się dopiero w latach 90. i był używany jedynie przez osoby zajmujące się bazami danych.

Gregory Piatetsky-Shapiro jest autorem terminu „Knowledge Discovery in Databases”, czyli odkrywania wiedzy ukrytej w bazach danych, który obecnie często stosuje się zamiennie z data miningiem. Ten drugi termin stał się popularny zwłaszcza w prasie, niemniej oba dotyczą w zasadzie tego samego. Data mining nie jest nową koncepcją. Od dawna jest obecny na rynku w wielu praktycznych zastosowaniach. Niemniej dopiero od paru lat zyskuje na popularności, na co wpływ ma zmieniające się



Wykorzystanie przestrzeni roboczej Statistica Data Miner w skoringu kredytowym.

otoczenie i dynamiczny rozwój technologiczny. Nie tylko komputery zapewniają coraz większą moc obliczeniową, dzięki czemu są w stanie szybciej przetwarzać cyfrowe dane. Również ludzie generują

ich coraz więcej. Dla zobrazowania: w 2003 r. całkowita ilość danych utworzonych od początku epoki cyfrowej wynosiła pięć eksabajtów, podczas gdy w 2014 r. pięć eksabajtów danych przybywało co dwa dni.

Ponieważ obecnie szybko rozwija się internet rzeczy, czyli pojawia się coraz więcej urządzeń łączących się z internetem i generujących dane, a także coraz więcej ludzi korzysta z komputerów czy smartfonów, w najbliższych latach czeka nas prawdziwa eksplozja danych. Firmy zajmujące się pamięciami masowymi +

W biznesie eksploracja danych to przede wszystkim docieranie do potencjalnych klientów szybciej niż konkurencja, ponieważ dzięki analizom danych wiemy, komu dokładnie i jaki produkt lub usługę powinniśmy zaoferować, a ponadto możemy bardziej precyzyjnie zaadresować naszą ofertę.

W CZYM SPRAWDZA SIĘ DATA MINING

Lista branż i tematów, w których eksploracja danych może przynieść wymierne korzyści

| | | |
|---|--|---|
| Handel detaliczny <ul style="list-style-type: none"> Segmentacja klientów Modelowanie odpowiedzi Rekomendowanie następnego prawdopodobnego produktu Profilowanie klientów o najwyższej wartości | Bankowość <ul style="list-style-type: none"> Ocena wiarygodności kredytowej Prawdopodobieństwo niespłacania zobowiązań Rentowność klienta Targetowanie klientów | Szkolnictwo wyższe <ul style="list-style-type: none"> Darowizny od absolwentów Pozyskiwanie studentów Zatrzymanie studentów Identyfikowanie zagrożonych studentów |
| Nauki przyrodnicze <ul style="list-style-type: none"> Opracowywanie leków i interakcje Typowe czynniki występujące u pacjentów chorych/zdrowych Klasyfikacja komórek rakowych Nadzór nad bezpieczeństwem leków | Ubezpieczenia <ul style="list-style-type: none"> Identyfikacja czynników ryzyka Próby oszustwa we wnioskach o odszkodowanie Analiza wartości klienta w czasie Łączenie polis Zatrzymanie pracowników | Służba zdrowia <ul style="list-style-type: none"> Rekomendowanie procedur dotyczących klientów Przewidywanie wyników pacjentów Wykrywanie oszustw Analiza notatek sporządzanych przez lekarzy i pielęgniarki |
| Telekomunikacja <ul style="list-style-type: none"> Utrata klientów Identyfikowanie możliwości sprzedaży związanej Wykrywanie włamań do sieci | Produkcja przemysłowa <ul style="list-style-type: none"> Analiza przyczyn błędów Analiza gwarancji Analiza niezawodności Analiza wydajności | Sektor publiczny <ul style="list-style-type: none"> Odchylenia i oszustwa podatkowe Analiza przestępczości Rozpoznawanie wzorców w nadzorze wojskowym |
| Motoryzacja <ul style="list-style-type: none"> Grupowanie opcji w pakiety dla poszczególnych segmentów klientów Analiza jakości dostawców Diagnozowanie problemów | Przemysł chemiczny <ul style="list-style-type: none"> Wykrywanie nowych związków chemicznych Łączenie molekuł w klastry Analiza wydajności produktów | Usługi komunalne <ul style="list-style-type: none"> Przewidywanie awarii linii energetycznych/sprzętu Łączenie produktów w pakiety Wykrywanie oszustw po stronie konsumentów |

+ obawiają się nawet, że nie będą w stanie nadać za popytem na przechowywanie danych, dlatego pracują nad nowymi technologiami, począwszy od zapisywania danych w nici DNA, po technologię HAMR, czyli magnetyczny zapis wspomaganym ciepłem. Ta druga ma zadebiutować w postaci gotowych nośników danych już w przyszłym roku.

Jeśli dodamy, że w 2010 r. ok. 1,8 mld ludzi miało dostęp do internetu, a do 2020 ma być aż 5 mld internautów, nie powinny dziwić prognozy ekspertów i analityków rynku, według których data mining może odmienić wiele branż, zapewniając firmom przetrwanie w szybko zmieniającej się rzeczywistości. – Ciekawe jest, że podczas gdy dawniej celem większości naszych projektów analitycznych i wdrożeniowych data mining była redukcja kosztów, obecnie coraz częściej pomagamy naszym klientom zwiększać ich przychody – mówi Jacek Jakubowski, prezes zarządu StatSoft Polska, producenta rozwiązań do wydobywania z danych użytecznej

wiedzy. – Obserwujemy również ostatnio wzrost liczby wdrożeń data mining w zastosowaniach technologicznych i produkcyjnych.

> DECYZJE PODPARTE DANymi

Na data mining uwagę powinny zwrócić przede wszystkim firmy, które mają bardzo dużo danych i mogą wykorzystać ich analizę do optymalizacji decyzji biznesowych. Drażnienie danych pozwala znaleźć więcej powiązań, często inaczej wręcz niemożliwych do odkrycia, pomiędzy posiadanymi informacjami. Za sprawą data

miningu można obniżyć koszty, zyskać większy zwrot z inwestycji z prowadzonych działań, a nawet pozyskać nowych klientów. – To doskonały sposób na zdobycie przewagi nad konkurencją – zaznacza Agnieszka Gawecka, marketing analytics lead z Microsoft Polska.

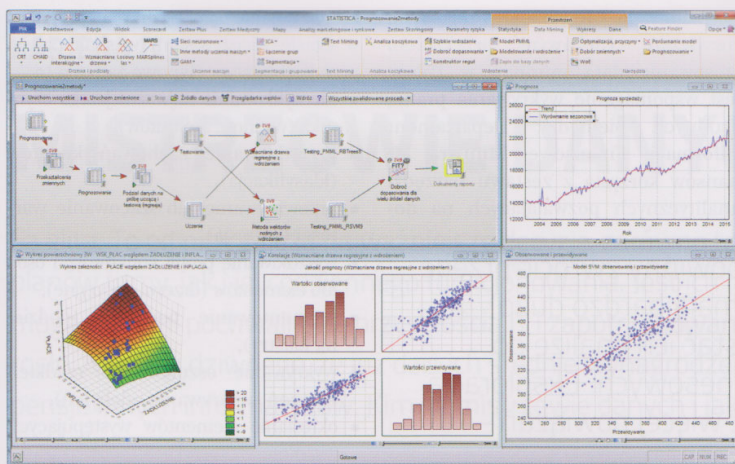
Rozwiązania data mining wykorzystywane są w bankowości, instytucjach finansowych, służbie zdrowia oraz firmach z sektora e-commerce. Korzystają z nich działy zajmujące się marketingiem, sprzedażą czy produkcją. Ponadto data mining sprawdza się w obszarach naukowych, gdzie ilości zbieranych informacji są gigantyczne.

W ostatnich latach dostrzec można jednak, że na rynku pojawia się coraz więcej start-upów, które oferują usługi drażnienia danych do innych celów, często nawet rozrywkowych. Dla przykładu amerykański start-up **Small Demons** analizuje książki, aby zaoferować ich czytelnikom wszystkie szczegóły zawarte przez autora. Można w ten sposób poznać nowe miejsca warte odwiedzenia, filmy warte obejrzenia czy też muzykę do posłuchania, a także wszelkie zależności i połączenia pomiędzy innymi książkami i bohaterami. Inny start-up – **Priceonomics** może analizować przeróżne branże i zjawiska, aby np. dowiedzieć się, jak szybko rozwija się dany produkt, kiedy jest największa szansa na obniżki w danym sklepie czy też ile recenzji otrzymała jakaś usługa w danym czasie. **ScrapHero** to z kolei start-up, który po wpisaniu adresu URL podpowie, jakie dane jest w stanie z niego wyciągnąć i w jakiej cenie.

DATA MINING W POLSCE

Poza bardzo popularną usługą data miningu świadczoną przez StatSoft (należącą obecnie do Dell) w Polsce działają również inne firmy specjalizujące się w eksploracji danych. Są to zarówno duże korporacje (oddziały Oracle, SAS), jak i mniejsze podmioty, takie jak Adept (adept.com.pl, Poznań) czy Dolnośląskie Centrum Analizy Danych (dcad.com.pl, Wrocław).

Warto również skorzystać z kursu online na temat data miningu, w którym poznamy wszystkie podstawowe informacje i porady: tinyurl.com/coursea-dtm.



Wyniki budowy modeli predycyjnych z wykorzystaniem Statistica Data Miner.

Kolejny przykład: **Kaggle** wykorzystuje drążenie danych do przeróżnych celów i udostępnia ciekawe poradniki, np. jak przewidzieć, którzy ludzie mogli przeżyć katastrofę Titanica poprzez przeanalizowanie całego wypadku z wykorzystaniem różnych narzędzi i języka programowania Python.

Przykłady te pokazują, że data mining wchodzi już w fazę demokratyzacji i moc obliczeniowa oraz wielkie zbiory danych są wykorzystywane do coraz bardziej oryginalnych celów.

W biznesie eksploracja danych to przede wszystkim docieranie do

potencjalnych klientów szybciej niż konkurencja, ponieważ dzięki analizom danych wiemy, komu dokładnie i jaki produkt lub usługę powinniśmy zaoferować, a ponadto możemy bardziej precyzyjnie zaadresować naszą ofertę. Dla przykładu Microsoft jest jedną z firm, które korzystają z data miningu w wielu obszarach, m.in. w planowaniu marketingu i sprzedaży, analizowaniu efektów prowadzonych działań czy też do analizy klientów oraz rynku. Pozwala to znacznie dokładniej oszacować potencjał sprzedaży, podnosić efektywność działań, dzięki tego typu

rozwiązaniom wiemy znacznie więcej o klientach, rynku i branży. Co więcej, nie tylko korporacje mogą sobie pozwolić na korzystanie z data miningu. Obecnie coraz częściej, dzięki spadkowi cen, usługi te dostępne są dla firm z sektora MŚP. Niemniej jednak data mining jest i będzie wykorzystywany przede wszystkim tam, gdzie daje największe zyski i gdzie wiedza o klientach, procesach biznesowych czy produkcyjnych jest zamieniana na konkretne korzyści, np. rozwiązanie kluczowego dla organizacji problemu.

> NOWE SZANSE BIZNESOWE

Twitter w marcu br. poinformował, że będzie sprzedawał biliony tweetów (wiadomości z tego serwisu społecznościowego) firmom, które chcą je analizować – są to zarówno komercyjne przedsiębiorstwa, jak i uczelnie, a nawet policja. Choć jak dotąd sprzedawanie danych nie miało dużego udziału w wynikach sprzedaży – wynosiło 70 mln dolarów z 1,3 mld dolarów w ubiegłym roku, Twitter chce to zmienić. Liczba potencjalnych klientów jest duża, są też firmy, które udostępniają odpowiednie narzędzia do drążenia tweetów. Przykładowo IBM również w marcu br. zaprezentował swoją pierwszą usługę działającą w chmurze, która udostępnia narzędzia do analizowania i eksploracji danych z Twittera.

ADWERTORIAL

Bitdefender GravityZone

W ubiegłym miesiącu (maj 2015 r.) miała miejsce premiera najnowszej generacji oprogramowania antywirusowego dla firm – Bitdefender GravityZone.

Aplikacja bazuje na specjalnie stworzonym neuronowym systemie ochrony sieci komputerowej BRAIN (Bitdefender Reflective Artificial Intelligence Networks), który nazywany jest antywirusową sztuczną inteligencją. System ten przetwarza prawie 200 tysięcy zapytań na sekundę od użytkowników na całym świecie, wyodrębniając wzorce złośliwego oprogramowania, i już po upływie trzech sekund

od ich publikacji jest w stanie zatrzymać związane z nimi zagrożenia.

Najnowsza wersja oprogramowania Bitdefender GravityZone zostaje rozszerzona o dodatkowe moduły:

- **Kontrola urządzeń** – ma pomóc w zapobieganiu infekcjom komputerów w rezultacie podłączenia do nich nieautoryzowanych urządzeń pamięci przenośnej itp. Administrator może stworzyć listę zaufanych urządzeń, które będą mogły się połączyć z komputerem;
- **Power User** – daje możliwość nadania uprawnień na poziomie administratora

wybranych użytkownikom końcowym.

Funkcja ta może znaleźć zastosowanie szczególnie w przypadku firm, które mają wiele lokalizacji. Nadane uprawnienia pozwalają na dostęp do punktów końcowych i modyfikację ustawień zabezpieczeń za pomocą lokalnej konsoli;

- **Konsola do centralnego zarządzania** – jest teraz bardziej funkcjonalna i przejrzysta w stosunku do konsoli znanej z poprzednich wersji.

Klienci posiadający wcześniejsze wersje oprogramowania Bitdefender mogą skorzystać z bezpłatnej aktualizacji.

Więcej informacji na bitdefender.pl/gz oraz kontakt@marken.com.pl

✦ W Polsce rynek data miningu nie jest w tak zaawansowanej fazie jak za granicą – Jesteśmy rozwijającym się, ale relatywnie biednym krajem. To naturalne, że dużych projektów i implementacji data mining jest mniej niż w wielu innych krajach – mówi Jacek Jakubowski. Ale jeśli tylko projekty

potrzebny do korzystania z tego typu rozwiązań. Ciekawym przykładem stosowania data miningu w Polsce jest wspomaganie wykrywania prania brudnych pieniędzy. Departament Informacji Finansowej Ministra Finansów korzysta z zestawu narzędzi analitycznych pozwalających na do-

które dostarczają wyniki i prognozy w czasie rzeczywistym w skali całego przedsiębiorstwa. Typowe zadania to:

- identyfikowanie najważniejszego czynnika (ważność atrybutów),
- przewidywanie zachowań klientów (klasyfikacja),
- przewidywanie lub szacowanie wartości (regresja),
- wyszukiwanie profili konkretnych osób lub elementów (drzewa decyzyjne),
- segmentowanie populacji (podział na klastry),
- wykrywanie oszustw lub rzadkich zdarzeń (wykrywanie odchyleń),
- określanie elementów występujących wspólnie w „koszykach” (powiązania).

Obecnie, kiedy przeżywamy prawdziwą eksplozję danych, potrzebne są bardziej rozbudowane techniki ich eksploracji. Mając dostęp do danych na temat np. 30 lub 40 mln klientów, wiedza o tym, że 2 mln z nich mieszka w jednym miejscu, nie jest wystarczająca. Chcemy znać inne szczegóły, takie jak: grupa wiekowa, średnie zarobki, często występujące potrzeby itp. Przez wiele lat w Polsce i na świecie łatwo można było poprawić efektywność organizacji i konkurencyjność rynkową, korzystając ze stosunkowo prostych działań organizacyjnych i technologicznych. Ale dziś wiedza o kliencie, organizacji oraz procesach jest coraz ważniejsza i to ona pozwala lepiej dopasować ofertę do klienta, zwiększyć sprzedaż czy ograniczyć ryzyko. Co więcej, dzięki drażeniu danych nie musimy podejmować prób odgadnięcia tego, co przyniesie przyszłość. Mamy dostęp do konkretnych danych i faktów. To właśnie data mining, wiedza wydobywana dzięki narzędziom analitycznym i wdrażana w postaci rozwiązań informatycznych, daje dzisiaj przewagę konkurencyjną. **IT**

Na data mining uwagę powinny zwrócić przede wszystkim firmy, które mają bardzo dużo danych i mogą wykorzystać ich analizę do optymalizacji decyzji biznesowych. Eksploracja danych pozwala znaleźć więcej powiązań – często inaczej wręcz niemożliwych do odkrycia – pomiędzy posiadanymi informacjami.

data mining są prowadzone w naszym kraju, efekty i zwrot z inwestycji nie odbiegają od tych z innych państw. Analizy ryzyka kredytowego i finansowego, przewidywanie awarii urządzeń i optymalizacji jakości produkcji, prognozowanie sprzedaży, przewidywanie zapotrzebowania na energię, wykrywanie nadużyć – tak właśnie przedstawiają się popularne zastosowania data miningu w Polsce.

Statistica Data Miner to popularny zestaw narzędzi data miningu, który został zaprojektowany w taki sposób, aby zapewniać łatwe i szybkie wykonywanie analiz danych. Oprogramowanie to oferuje przejrzysty interfejs, włącznie z kreatorem analiz, który prowadzi użytkownika krok po kroku przez wszystkie etapy wydobywania wiedzy z danych. PKO BP korzysta np. z tego oprogramowania do oceny ryzyka i budowy modeli scoringowych, KDPW stosuje data mining przy prognozowaniu wielu parametrów, głównie dotyczących rynku papierów wartościowych, natomiast PKN Orlen używa data miningu m.in. do badania efektywności promocji i ich wpływu na zachowania klientów. W Polsce na data miningu polegają głównie duże firmy i instytucje, które mają odpowiednich analityków i know-how

głębną analizę danych o transakcjach, poszukiwanie wzorców i reguł występujących w danych i na tej podstawie typuje podejrzane transakcje.

Firmy oferujące usługi data miningu prowadzą też odpowiednie szkolenia, przez co zdobycie wiedzy z zakresu eksploracji danych nie jest już tak trudne jak dawniej.

> WYKRYĆ POTRZEBY, ZANIM SIĘ POJAWIĄ

Mając więcej informacji niż konkurencja, znacznie szybciej możemy przewidywać zmiany rynkowe i reagować na nie oraz dostosowywać się do nowych trendów. I faktycznie, data mining doskonale sprawdza się w przewidywaniu zmian rynkowych i dzięki niemu można tworzyć skuteczne modele predykcyjne. Te mogą przewidywać prawdopodobieństwo wystąpienia dowolnego zjawiska, np. zaniechania spłat, wystąpienia wypadku, odejścia klienta czy dokonania zakupu.

Dla przykładu firma Oracle koncentruje się na świadczeniu usług data miningu i dzięki nim analitycy oraz twórcy aplikacji mogą używać interfejsów API SQL do tworzenia aplikacji, które automatycznie eksplorują dane w topologii gwiazdy. Ma to na celu tworzenie i wdrażanie modeli predykcyjnych,

Autor jest dziennikarzem specjalizującym się w nowych technologiach i biznesie online. Od 10 lat publikuje w tytułach prasowych i w internecie. Świadczy również usługi konsultacyjne start-upom z branży IT.