

ROZDZIAŁ 3

Systemy *Business Intelligence* (BI)

3.1. Wprowadzenie

Business Intelligence to bardziej pojęcie marketingowe niż naukowe czy techniczne¹. Zostało wprowadzone i propagowane na przełomie lat 80. i 90. XX w. przez analityków międzynarodowego rynku ICT, pracujących dla organizacji międzynarodowych, takich jak Gartner Group, Forrester Research czy IDC, na określenie architektury informatycznej, opartej na zbiorze aplikacji transakcyjnych i systemach wspomagania decyzji mających na celu łatwy dostęp do informacji zarządczej pochodzącej z różnych rozproszonych źródeł [Wolny, 2004]. Za twórcę pojęcia *Business Intelligence* uznawany jest Howard Dresner, amerykański informatyk związany obecnie z firmą badawczo-analityczną Dresner Advisory Services, LLC, a w przeszłości m.in. z Gartner Group².

- 1 *Business Intelligence* najczęściej nie jest tłumaczone na język polski, choć są opracowania, w których używane są nazwy „inteligencja biznesowa” (co uznaje się za tzw. kalkę językową), „system informacji zarządczej”, „system wywiadu gospodarczego”, „system wiedzy biznesowej” etc.
- 2 Dresner posłużył się pojęciem *Business Intelligence*, które było używane w informatyce od co najmniej lat 50. ubiegłego wieku. W 1958 r. na międzynarodowej konferencji International Conference on Scientific Information (ICSI) zorganizowanej w Waszyngtonie, H.P. Luhn z International Business Machines (IBM) oraz H.M. Ohlman z System Development Corporation (SDC) zaprezentowali systemy do automatycznego indeksowania dokumentów tekstowych. Luhn nazwał swój system „indeksem słów kluczowych” – Keyword-in-Context Index (KWIC), natomiast Ohlman – „indeksem permutacyjnym” – Permutation Indexing. Niemalże w tym samym czasie Luhn przedstawił system „selektywnej dystrybucji informacji” – Selective Dissemination of Information (SDI) jako jeden z elementów szerszego pojęcia *Business Intelligence* system (BI). SDI zakładał użycie komputera do wyboru istotnych dla użytkownika dokumentów spośród większej ich liczby. H.P. Luhn zdefiniował biznes jako zbiór działań celowych. W tym sensie biznes jest obecny w nauce, technologii, handlu, przemyśle, działaniach prawnych, administracji, działaniach związanych z obronnością etc. Inteligencję postrzegał jako umiejętność uchwycenia sensu wzajemnych powiązań faktów w sposób umożliwiający kierowanie działań na rzecz pożądanego celu. System BI definiował jako automatyczny system informatyczny rozwijany w celu selektywnej

Howard Dresner zwrócił uwagę na to, że ówczesne koncepcje i metody poprawy efektywności podejmowania decyzji biznesowych oraz poprawy wydajności procesów biznesowych wymagały użycia systemów informatycznych bazujących na faktach i że fakty te tworzone były w wyniku analiz przeprowadzanych na danych rozproszonych w wielu systemach komputerowych, a w związku z tym wymagających oczyszczenia, integracji, agregacji i transformacji. Pojawienie się *Business Intelligence* miało swoje podłoże w równoległe dziejących się procesach zmian przedsiębiorstw dotyczących wzrostu znaczenia zasobów niematerialnych. Mniej więcej w tym samym okresie, w którym kształtowała się idea *Business Intelligence*, rozwijała się koncepcja piramidy zasobów informacyjnych (*DIKW pyramid – Data-to-Information-to-Knowledge-to-Wisdom*) za sprawą m.in. Russella Ackoffa [Ackoff, 1989]. Robert Kaplan i David Norton opracowali zrównoważoną kartę wyników [Kaplan, 1996] oraz mapę strategii [Kaplan, 2001], które do oceny efektywności procesów biznesowych wykorzystywały wskaźniki odwołujące się do potrójnego salda wyników (*Triple Bottom Line*) [Elkington, 2018], a więc wskaźniki ekonomiczne, społeczne i środowiskowe, wreszcie w tym także czasie ukształtowały się koncepcje kluczowych wskaźników efektywności (*Key Performance Indicators, KPI*) w formie, którą znamy do dzisiaj, i które można było postrzegać z poziomu *Business Intelligence* jako fakty. W chwili obecnej pojęcie *Business Intelligence* zastępowane jest często pojęciem *Business Analytics*. W naszym podręczniku pozostaniemy przy tradycyjnej nazwie *Business Intelligence*.

Business Intelligence umożliwia proste tworzenie długookresowej strategii funkcjonowania systemu informacji zarządczej w organizacji na podstawie pewnego rodzaju szkieletu do budowy systemu informatycznego (*framework*), czyli zestawu gotowych komponentów do budowy komputerowego systemu informacji zarządczej, na którym można zbudować rozwiązanie dopasowane do własnych potrzeb, uwarunkowań projektowych i środowiskowych, spełniające potrzeby i oczekiwania różnych interesariuszy [Ślęzak, 2016]. *Framework* do budowy systemów klasy *Business Intelligence* obejmuje następujące elementy:

dystrybucji informacji do różnych jednostek organizacji przemysłowych, naukowych i rządowych. System BI miał korzystać z urzędów do automatycznego tworzenia abstraktów dokumentów i automatycznego kodowania dokumentów w celu stworzenia profili zainteresowań dla każdego z „punktów decyzyjnych” w organizacji. System BI realizować miał zatem usługi informacyjne w sensie zarządzania dokumentami oraz dostarczania odpowiednich informacji w punktach decyzyjnych, utożsamianych z indywidualnymi osobami, grupami lub całą organizacją [Gontar, 2016].

1. Ramowy zestaw wskaźników efektywności (KPI) obejmujący także ramy analityczne umożliwiające wypracowanie, na potrzeby systemu raportowania, narzędzi do analizy porównawczej dla KPI (*benchmarking*³) oraz do interpretacji wartości KPI – ustalenia poziomu efektywności procesów w zależności od wartości wskaźników;
2. System analiz decyzyjnych bazujący na statystyce, uczeniu maszynowym i sztucznej inteligencji, który korzystałby ze wspomnianego zestawu wskaźników, a więc współczesna wersja systemu wspomaganego decyzji (*Decision Support System, DSS*);
3. System raportowania obejmujący wizualizację danych, systemy informowania kierownictwa i systemy OLAP (*On-line Analytical Processing*), który pozwalałby na obliczanie wskaźników z częstotliwością, która jest wymagana dla określonego obszaru, „rozwijanie” wskaźników w dół (tj. w podziale na kategorie produktów, lokalizacje, kanały sprzedaży), porównywanie wewnątrz poszczególnych kategorii i w odniesieniu do konkurencji, obserwacje zmian w czasie;
4. Platformy wspomagające pracę zespołową. Dobrym przykładem wyjaśniającym ideę tego typu platformy jest tworzona obecnie Polska Platforma Przemysłu 4.0, której zadaniem jest rozpowszechnianie wiedzy, umiejętności oraz standardów technicznych i zarządczych, które są fundamentem udanej transformacji do Przemysłu 4.0, ale także wspólna praca nad dokumentami, wspólne projekty, prezentacje zespołowe etc.;
5. Systemy zarządzania wiedzą. Umożliwiają tworzenie, upowszechnianie i wykorzystanie wiedzy o stanie przedsiębiorstwa do realizacji celów strategicznych organizacji.

W zasadzie możliwe jest opracowanie ramowego zestawu wskaźników efektywności dla każdej branży, uwzględniającego jej specyfikę. W naszej książce zaprezentujemy pokrótce propozycję *Consortium for Advanced Management-International (CAM-I)* dotyczącą generycznego modelu organizacji⁴, tzw. *Metrics Reference Model (MRM)*. Ważną rolę w wykorzystaniu mierników efektywności do określania stanu przedsiębiorstwa odgrywa analiza wielowymiarowa opracowana w roku 1996 przez Roberta S. Kaplana i Davida P. Nortona pod nazwą

3 *Benchmarking* jako metoda wprowadzania zmian w organizacji zasadza się na idei mierzenia wyrobów, usług i procedur względem najsilniejszych konkurentów lub tych firm, które uznawane są za liderów przemysłu (www.pi.gov.pl). CAM-I w swoim modelu referencyjnym podaje wartości wskaźników strategicznych dla modelowej organizacji, do której można porównać wskaźniki osiągnięte przez naszą organizację.

4 Model generyczny – inaczej: ogólny, możliwy do zastosowania w każdej branży, pozwalający na standaryzację metod oceny efektywności organizacji.

zrównoważonej karty wyników [Kaplan, 1996]. Jest ona stosowana jako system pomiarowo-analityczny obejmujący podstawowe cele strategiczne organizacji i mierniki ich realizacji, przechowywane w kilku perspektywach⁵. Perspektywa finansowa dotyczy wartości organizacji i obejmuje różne mierniki rachunkowości, perspektywa klienta odnosi się do grup klientów i mierników ich satysfakcji, lojalności etc., perspektywa procesów wewnętrznych dotyczy jakości procesów, wreszcie perspektywa wzrostu i uczenia się odnosi się do instrumentów dyfuzji wiedzy w organizacji. Możliwe jest także zastosowanie *Business Model Canvas* z 2010 roku, którego autorami są Alex Osterwalder i Yves Pigneur [Osterwalder, 2012]. Należy pamiętać o tym, że mierniki oceny efektywności procesów biznesowych ewoluują wraz z rozwojem celów strategicznych organizacji, ale także wraz z rozwojem teorii zarządzania strategicznego. Możemy zatem przyjąć jako punkt startu jeden z proponowanych „standardów” albo na bazie własnego doświadczenia i dobrych praktyk opracować oryginalny zestaw wskaźników.

Platformy wspomagające pracę zespołową umożliwiają współdzielenie zasobów (tych wymienionych w poprzednich punktach: system wskaźników efektywności, system analiz decyzyjnych, system raportowania), mobilny dostęp do pulpitów menedżerskich, dzielenie się komentarzami z każdego z pulpitów sterujących. Możemy tu odwołać się do koncepcji zwinnego BI (*Agile BI*) [Ambler, 2011], opartego na Manifeście zwinnego wytwarzania oprogramowania będącego deklaracją wspólnych zasad dla zwinnych metodyk tworzenia oprogramowania, który został opracowany w roku 2001 i zakładał odejście od metodyki tradycyjnej na rzecz wspierania pracy z ludźmi i ich wzajemnych interakcji, wykorzystania w tym celu stosownego oprogramowania, współpracy z klientem, reagowania na zmiany [Beck, 2001]. W tym kontekście ważne jest wykorzystanie platform wspomagających pracę zespołową w zakresie:

1. Wsparcia technologicznego dla komunikacji i pracy grupowej w obrębie zespołu projektowego tworzącego rozwiązania BI oraz pomiędzy zespołem projektowym a biznesem, w tym w zakresie urządzeń mobilnych;
2. Propagowania wśród biznesu informacji o relacji zwinne BI/zwinna (elastyczna) organizacja, o możliwościach do przeprowadzenia analizach, o interpretacji wyników badań, metodykach zwinnych w zarządzaniu projektami (np. *scrum*, *kanban*) etc.;

5 Tradycyjnie, zrównoważona karta wyników dotyczy perspektywy finansowej, perspektywy klienta, perspektywy procesów wewnętrznych oraz perspektywy wzrostu i uczenia się; w kontekście strategii zrównoważonego rozwoju konieczne jest ich zmodyfikowanie.

3. Propagowania w obrębie zespołu projektowego oraz pomiędzy zespołem projektowym a biznesem nowych idei związanych z pojęciem sukcesu projektu; odejście od tradycyjnego „ukończenia na czas, po oczekiwanych kosztach, z mechanizmami i funkcjami zgodnymi z wyjściową specyfikacją” na rzecz użyteczności dla biznesu i dostarczenia wartości;
4. Konieczności ścisłej współpracy pomiędzy zespołem projektowym a biznesem w zakresie budowy aplikacji BI i jej utrzymania, w tym ciągłych zmian w funkcjonalności aplikacji BI wynikających z podtrzymania wysokiej elastyczności (zwinności) organizacji i ciągłego doskonalenia procesów biznesowych.

Zazwyczaj strategia funkcjonowania systemu informacji zarządzanej w organizacji budowana jest na rozwiązaniach komercyjnych dostawców, jak *Comarch Business Intelligence*, *Samsung Business Consulting*, *Oracle Business Intelligence*, *Hogart Business Intelligence*, *Power Business Intelligence*, *SAP Business Intelligence*, *SAS Business Intelligence*, *Tableau* etc. Możliwe jest jednak zbudowanie takiej wizji na podstawie rozwiązań niekomercyjnych, jak *Business Intelligence and Reporting Tools (BIRT)*, *MicroStrategy Desktop Business Intelligence* etc.⁶

Business Intelligence wpisuje się w tradycyjny model DIKW (*Data-to-Information-to-Knowledge-to-Wisdom*), nazywany też piramidą zasobów informacyjnych (danych, informacji, wiedzy i mądrości) organizacji (rys. 3.1).

U podstaw modelu DIKW leżą pomiary ilościowe zmian zachodzących w środowisku biznesowym wywołanych przez procesy biznesowe, realizacją transakcji etc. Wynikiem są dane o transakcjach, dane o procesach etc.

3.2. Podstawowe pojęcia

3.2.1. Proces zarządzania efektywnością i łańcuch poznawczy dane-informacja-wiedza-mądrość (*Data-to-Information-to-Knowledge-to-Wisdom*, DIKW)

W celu określenia, w jaki sposób systemy informatyczne mogą wspierać decydentów w przedsiębiorstwie, nie wystarczy zrozumienie samego zadania wspomaganie decyzji, co będzie przedmiotem rozważań

6 W dalszej części rozdziału będzie mowa o rozwiązaniach komercyjnych i językach zapytań wywodzących się z języka SQL: MDX, DAX, MDX.

w II części podręcznika. Podejmowanie uzasadnionych decyzji wymaga bowiem stałego monitorowania sytuacji w przedsiębiorstwie i monitorowania efektów własnych działań (monitorowania wydajności podejmowanych działań/monitorowania realizacji celów) poprzez system pomiaru efektywności i dostęp do wiedzy o decyzjach podejmowanych w przeszłości (poprzez system zarządzania wiedzą), co zmniejsza potrzeby odkrywania wiedzy na nowo. Wyjaśnimy to, wykorzystując łańcuch DIKW, który pozwala na zrozumienie relacji między danymi, informacją, wiedzą i mądrością strategiczną przedsiębiorstwa [Kisielnicki, 2008].

Systemy BI pozwalają tworzyć systemy zarządzania efektywnością na podstawie kluczowych wskaźników efektywności (*Key Performance Indicators, KPI*).

Rys. 3.1. Etapy procesów podejmowania decyzji i rozwiązywania problemu



Źródło: opracowanie własne.

Jak widzimy na rysunku 3.1, każde z kolejnych ogniw łańcucha zależy od pozostałych.

W pierwszym ogniwie łańcucha, na etapie przetwarzania danych, poszukiwana jest odpowiedź na pytanie, w jaki sposób ocenić efekty własnych działań i jakie dane wykorzystywane są do pomiaru efektywności. Zbierana jest informacja odnosząca się do zakresu problemu zarządzania efektywnością. Wypracowanie własnych procedur działania

w tym zakresie jest często poprzedzone zastosowaniem jednolitych standardów. Najbardziej powszechne jest wykorzystanie kluczowych wskaźników efektywności (*Key Performance Indicies*) oraz zastosowanie modelu referencyjnego metryk efektywności (*Metrics Reference Model*, MRM) opracowanego przez *Consortium for Advanced Management-International* (CAM-I), zrównoważonej karty wyników albo szablonu modelu biznesowego. Dane, które będą wspierać obliczanie wybranych KPI, pochodzą zazwyczaj z istniejących w organizacji relacyjnych baz danych, o których była mowa w rozdziale 2 o bazach danych.

W fazie przetwarzania informacji następuje ekstrakcja danych ze zidentyfikowanych źródeł danych, integracja i transformacja danych na potrzeby wyliczeń wskaźników efektywności i ładowanie przetworzonych danych do jednej hurtowni danych lub wielu hurtowni tematycznych. Proces gromadzenia informacji realizowany jest w tradycyjnych systemach BI w trybie cyklicznym. W środowisku IoT możliwe jest także gromadzenie informacji w trybie ciągłym. Należy zauważyć, że aktualizacja zgromadzonej informacji polega na wprowadzeniu nowej informacji do hurtowni danych (hurtowni tematycznych). Informacja nie jest usuwana, ani zastępowana aktualną. Hurtownia danych jest zazwyczaj budowana jako baza relacyjna. Językiem wykorzystywanym do interpretacji informacji jest więc tradycyjnie SQL. Możliwe jest także zbudowanie wokół hurtowni danych bazy wielowymiarowej (bazy OLAP). W takiej sytuacji do interpretacji informacji wykorzystywane są inne języki, np. MDX czy DAX.

Etap przetwarzania wiedzy wymaga określenia toku akcji podejmowanych w celu przeanalizowania wskaźników efektywności i powiązania ich z celami strategicznymi organizacji. Zazwyczaj wykorzystywane są tu modele statystyczne, uczenia maszynowego i sztucznej inteligencji. Językiem wykorzystywanym do analizy informacji jest np. DMX, aczkolwiek istnieje wiele innych języków programowania mogących mieć zastosowanie w tej fazie.

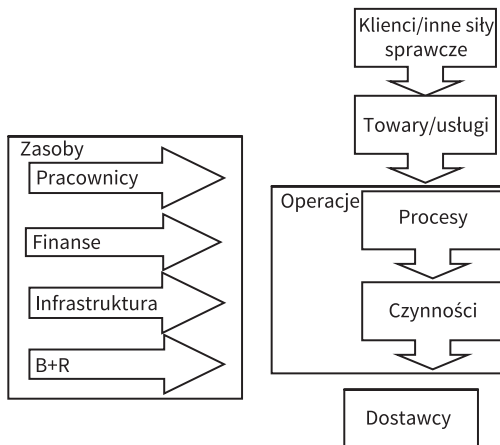
Na etapie korzystania ze zgromadzonej wiedzy (określanej jako mądrość strategiczna przedsiębiorstwa) następuje weryfikacja procesów/projektów biznesowych na podstawie zgromadzonej wiedzy z punktu widzenia strategii realizowanej przez organizację i odpowiednie działania dostosowane do stanu przedsiębiorstwa: np. reinyżeria procesów, *lean management*, *total quality management* czy metodologia sześć sigma, a w przypadku projektów – hierarchizacja i wybór projektów wspierających w najwyższym stopniu realizację celów strategicznych przedsiębiorstwa.

3.2.2. Model referencyjny dla kluczowych wskaźników efektywności

Model referencyjny (inaczej model odniesienia) może być postrzegany jako model wzorcowy. Stanowi często punkt wyjścia do prac nad doborem docelowych kluczowych wskaźników efektywności w organizacji. Jest opracowywany na podstawie dobrych praktyk, pozwala więc korzystać z doświadczeń innych firm. Przyspiesza prace nad opracowaniem wskaźników docelowych, dlatego ważne jest, aby przed przystąpieniem do prac nad budową własnego systemu BI wybrać stosowny model referencyjny: ogólny lub branżowy.

Wykorzystamy ogólny model referencyjny opracowany przez CAM-I⁷. Model nosi nazwę MRM i odwołuje się do generycznego modelu organizacji, a więc do modelu właściwego wszystkim organizacjom. W modelu MRM wyodrębniono kilka typowych elementów struktury organizacyjnej każdej typowej organizacji (rys. 3.2), dla których zaproponowano listę kluczowych wskaźników wraz z ich ogólną charakterystyką. Ustalono pożądane wartości, do których można odwołać się w analizie porównawczej (*benchmarking*) i wzajemne relacje między wskaźnikami. Scharakteryzujemy pokrótce model MRM, ponieważ modele referencyjne tego typu pełnią istotne funkcje we wspomagananiu procesu budowy systemu BI.

Rys. 3.2. Taksonomia struktury organizacyjnej generycznego modelu organizacji według CAM-I



Źródło: [Chaudhuri, 2010].

7 <http://www.cam-i.org> [dostęp: 27.05.2018].

W modelu MRM wyróżniono dziewięć elementów struktury organizacyjnej typowej organizacji (inaczej – dziewięć kategorii referencyjnych wskaźników efektywności):

1. Klienci (lub inne siły napędowe pobudzające rozwój biznesu);
2. Produkty i usługi;
3. Procesy i zdarzenia;
4. Pracownicy;
5. Finanse;
6. Badania i rozwój;
7. Infrastruktura;
8. Zasoby;
9. Dostawcy.

CAM-I nie zdecydowało się opublikować referencyjnych wskaźników efektywności dla kategorii „procesy i zdarzenia”, wychodząc z założenia, że te zależne są od natury procesu, nie istnieje zatem możliwość zdefiniowania wyczerpującej listy standardowych wskaźników dla „typowego” procesu. Dla pozostałych kategorii model MRM zawiera propozycje wskaźników zdefiniowanych w trzech perspektywach: kosztów, procesów i wyników oraz następujące informacje:

1. Rodzaj wskaźnika. Decyduje o strukturze wskaźnika i możliwości jego rozwijania w ramach określonej hierarchii;
2. Definicja/obliczenia. Opisany jest sposób obliczania na podstawie dostępnych danych;
3. Mocne strony/możliwości. Znane zalety korzystania z metryki;
4. Słabości/problemy/zagrożenia. Znane wady korzystania z metryki;
5. Ustawienie celu. Dyskusja o pożądanach wartościach wskaźnika;
6. Możliwe korelacje/koszyk metryk. Powiązania z innymi wskaźnikami, wskazanie siły związku (np. słaba, umiarkowana, mocna) i wskazanie na rodzaj zależności (np. korelacja ujemna, dodatnia).

W tabelach 3.1 i 3.2 – korzystając z [Chaudhuri, 2010] – podajemy przykładowe metryki z kategorii „klient” oraz „pracownik” w perspektywach wyników, kosztów i procesów.

Tabela 3.1. Metryki efektywności, kosztów i procesów z kategorii „klient”

| Perspektywa | Wyników | Procesów | Kosztów |
|-------------|----------------------------|--|----------------------------|
| Typ | Klient/satysfakcja klienta | Klient/satysfakcja klienta | Klient/satysfakcja klienta |
| Miara | Średnia ocena klientów | Wskaźnik rozstrzygniętych spraw klientów | Poziom reklamacji |

| Perspektywa | Wyników | Procesów | Kosztów |
|--------------------------------|---|---|---|
| Definicja/obliczenia | Średnia arytmetyczna ocen klientów | Odsetek rozwiązanych spraw | Wartość reklamacji względem wartości sprzedaży |
| Silne strony/ możliwości | Łatwy do uchwycenia za pomocą ankiet konsumenckich | Wsparcie dla badania satysfakcji klientów Rozwiązywanie problemów (reklamacji) na tym poziomie jest tańsze | Miara jakościowa (może nie dotyczyć wszystkich produktów) |
| Słabe strony/ problemy/ ryzyka | Satysfakcja jest miarą subiektywną. Wskaźniki odpowiedzi na ankiety są zazwyczaj bardzo niskie. Ekstremalne opinie dominują w ankietach, ponieważ klienci o takich opiniach są bardziej zmotywowani do wypełniania ankiet | W niektórych przypadkach może być trudno mieralne Subiektywność w pozytywnym rozstrzygnięciu reklamacji | Trudności w zróżnicowaniu |
| Ustawienia docelowe | Zwykle pożądane są wysokie wartości | Wysokie wartości | Niskie wartości |

Źródło: opracowanie własne na podstawie [Chaudhuri, 2010].

Tabela 3.2. Metryki wyników, kosztów i procesów z kategorii „pracownik”

| Perspektywa | Wyników | Procesów | Kosztów |
|--------------------------|---|--|---|
| Typ | Utrzymanie pracownika | Rekrutacyjny | Rekrutacyjny |
| Miara | Średni wskaźnik rotacji | Stosunek liczby wywiadów do liczby otwartych stanowisk | Średni całkowity koszt zatrudnienia |
| Definicja/obliczenia | Liczba pracowników odchodzących/liczba pracowników ogółem | Kandydaci, którzy przychodzą na rozmowę kwalifikacyjną na jedno stanowisko | Średnie koszty zmienne na zatrudnionego + średni czas pracy rekrutera i ankietera (* ich godzinowa pensja) na zatrudnionego |
| Silne strony/ możliwości | Łatwy do uchwycenia Łatwy do zrozumienia | Wskazuje siłę reklamy (precyzję opisu stanowiska pracy) Może wskazywać na szybkość skuteczności reakcji | Średni koszt zatrudnienia liczony na podstawie kosztów zmiennych stwarza trudności obliczeniowe |

| Perspektywa | Wyników | Procesów | Kosztów |
|--------------------------------------|--|--|--|
| Stabe strony/ problemy/ ryzyka | Wysokie koszty zatrzymania pracowników (urlopy etc.) | Nie rozróżnia naszego wyboru rozmówców i ich wyboru odnośnie do stanowiska pracy Nie wiadomo, jak ustawić właściwą wartość docelową Wysokie wartości mogą oznaczać wyższe koszty | Pełne koszty w porównaniu z kosztami zmiennymi trudności obliczeniowe Obniżenie tego wskaźnika stanowiłoby zachętę do rotacji (wzrost wolumenu do niższych średnich kosztów) |
| Ustawienia docelowe | 0% niekoniecznie jest dobrą wartością – zależy to od pracy | Poszukuje liczby mniejszej niż 1, aby podjąć działanie (zależy od wartości miary) | Zwykle pożądane są niskie wartości |

Źródło: opracowanie własne na podstawie [Chaudhuri, 2010].

Pierwotnie miary wykorzystywane w analizie efektywności funkcjonowania przedsiębiorstwa były zorientowane na tworzenie wartości przedsiębiorstwa oraz na cele strategiczne (zrównoważona karta wyników). Obecnie takie samo podejście można dostrzec w zarządzaniu przedsiębiorstwem na dowolnym poziomie jego struktury. Należy także zwrócić uwagę na następujące istotne kwestie:

1. W modelu generycznym mamy do czynienia z szeregiem wyróżnionych obszarów funkcjonowania przedsiębiorstwa i dla każdego z nich możliwe jest zdefiniowanie rozbudowanego zestawu wskaźników efektywnościowych, a co za tym idzie – szeregu rozbudowanych analiz. Istnieje więc pokusa budowy nie jednej rozbudowanej hurtowni danych dla całego przedsiębiorstwa, a wielu tematycznych hurtowni w określonych obszarach jego funkcjonowania, wokół których łatwiej jest zbudować systemy analityczne. Możliwe jest zatem opracowanie systemu wskaźników oraz zbudowanie w przedsiębiorstwie systemu BI typu CRM (*Customer Relationship Management*) dla kontaktów z klientami czy systemu BI typu SCM (*Supply Chain Management*) dla zarządzania łańcuchem logistycznym;
2. Model generyczny jest uproszczeniem, można go wprowadzić zastosować do dowolnej organizacji, niemniej istnieją takie obszary funkcjonowania przedsiębiorstw, dla których celowe jest zbudowanie zestawu wskaźników od zera;
3. BI dostarcza technologię i standaryzowane podejście (oparte na dobrych praktykach) do zbierania, raportowania i analizy wskaźników efektywności procesów biznesowych w sposób zautomatyzowany

na wszystkich poziomach organizacji. Nie oznacza to jednak, że bez systemu BI jest to niemożliwe. Można tworzyć ekstrakty danych, budować na ich podstawie stosowne wskaźniki w arkuszach kalkulacyjnych i przeprowadzać ich analizę, wykorzystując dowolne narzędzie statystyczne.

Istotnym elementem modelu referencyjnego MRM jest tabela zależności, która zawiera informację o naturze zależności między wskaźnikami (pozytywnej lub negatywnej korelacji) oraz jej sile (siła zależności korelacyjnej (+ dodatniej; – ujemnej) wyrażona jest nieprecyzyjnie jako duża (D), umiarkowana (U) lub słaba (S)) (tab. 3.3).

Tabela 3.3. Tabela zależności między wskaźnikami z kategorii „klient” i „pracownik”

| | Średnia ocena klientów | Stopa zwrotu z powodu wady lub zakończenia okresu próbnego wykorzystania | Wskaźnik rozstrzygniętych spraw klientów | Średni wskaźnik rotacji | Stosunek liczby wywiadów do liczby otwartych stanowisk | Średni koszt zatrudnienia |
|--|------------------------|--|--|-------------------------|--|---------------------------|
| Średnia ocena klientów | | U | U+ | S | S | S |
| Stopa zwrotu z powodu wady lub zakończenia okresu próbnego wykorzystania | U- | | U | S | S | S |
| Wskaźnik rozstrzygniętych spraw klientów | U+ | U | | S | S | S |
| Średni wskaźnik rotacji | S | S | S | | S | S |
| Stosunek liczby wywiadów do liczby otwartych stanowisk | S | S | S | S | | S |
| Średni koszt zatrudnienia | S | S | S | S | S | |

Źródło: opracowanie własne na podstawie [Chaudhuri, 2010].

3.2.3. Analiza i monitoring stanu przedsiębiorstwa

Ocena aktualnego stanu przedsiębiorstwa wymaga zastosowania obserwacji (monitoringu) środowiska biznesowego i prognozowania zmian w nim zachodzących. Najczęściej monitoring określany jest jako okre-

sowe pobieranie danych próbkowanych i ich automatyczną analizę. Raportowanie wydajności można więc ustalić dopiero po fakcie, przeprowadzając analizę off-line. Można jednak także wykorzystać systemy BI wbudowane w systemy wspierające zarządzanie procesowe, gdzie realizowane jest zbieranie, analizowanie i interpretowanie danych w czasie rzeczywistym. Przedmiotem analizy są tu strumienie zdarzeń. Koncepcja ta odwołuje się tym samym do zasad ciągłego doskonalenia procesów obecnych w naukach o zarządzaniu od lat 90. ubiegłego wieku i metod zarządzania jakością. Systemy BI to w tym sensie systemy monitorujące procesy (parametry systemu operacyjnego), raportujące kluczowe dla oceny wydajności procesu metryki, generujące alerty o stanie procesu, wszczynające natychmiastowe działania.

Przedmiotem analizy BI mogą być także procesy obejmujące relacje pomiędzy przedsiębiorstwami, klientami, dostawcami, akcjonariuszami etc., którzy razem tworzą sieć semantyczną zbudowaną na zasadzie sieci społecznych. W tej koncepcji nie jest możliwe opomiarowanie procesów za pomocą aparatury kontrolno-pomiarowej. W tego typu systemach BI procesy musiałyby być monitorowane przez autonomiczne agenty.

3.2.4. Benchmarking

Ocena aktualnego stanu przedsiębiorstwa polega na formułowaniu odpowiednich wskaźników efektywności, a następnie na określeniu i wyborze sposobu ich wykorzystania do porównań ze stanem wybranym jako punkt odniesienia, np. stanem przedsiębiorstwa sprzed roku, konkurencyjnymi przedsiębiorstwami, najlepszymi praktykami (wzorcowymi procesami realizowanymi wewnątrz i na zewnątrz przedsiębiorstwa). Tradycyjnie w przypadku analiz porównawczych BI możemy mówić o następujących kategoriach:

1. Stan przedsiębiorstwa z poprzedniego okresu;
2. Stan największych konkurentów;
3. Stan liderów branży.

Mówiąc o analizie porównawczej, należy zwrócić uwagę, że wymaga ona zgromadzenia informacji o branży, uczestnikach rynku etc., tzn. zastosowania konkurencyjnego wywiadu gospodarczego [Chmielewski, 2009]. Wywiad gospodarczy przedsiębiorstw to dziedzina nauk o zarządzaniu wywodząca się z państwowego wywiadu gospodarczego. Procesy wywiadu gospodarczego przedsiębiorstw obejmują gromadzenie i analizę danych wywiadowczych – pochodzących z legalnych źródeł wewnętrznych i zewnętrznych – oraz przekazywanie raportów

wywiadowczych decydyntom w celu poprawy efektywności krytycznych procesów biznesowych przedsiębiorstwa. Pierwsze jednostki wywiadu gospodarczego tworzone były w korporacjach branży ICT: MOTOROLA, IBM, HP, Intel, począwszy od lat 80. XX wieku. Wywiad gospodarczy przedsiębiorstw wykorzystuje narzędzia BI do gromadzenia, wyszukiwania i analizowania danych wywiadowczych. Koncepcja BI obejmuje dwa zagadnienia: kulturę zarządzania oraz narzędzia teleinformatyczne do zarządzania informacjami i ich analizowania. Celem wywiadu gospodarczego jest określenie, jakie informacje są potrzebne, w jaki sposób powinny zostać zebrane, jak powinny być zorganizowane, gdzie powinny być przechowywane oraz kto w przedsiębiorstwie powinien mieć do nich dostęp. Analizowanie informacji to dziedzina analityki biznesowej. W zakresie kultury zarządzania istotna jest organizacja jednostki przedsiębiorstwa realizującej procesy BI na potrzeby wywiadu gospodarczego. Gärtner Group zaproponował na początku XXI wieku koncepcję centrum kompetencyjnego BI – interdyscyplinarnego zespołu o trwałej, formalnej strukturze organizacyjnej, będącego własnością przedsiębiorstwa, z którego rekrutują się jego pracownicy, przed którym stawia się określone zadania, definiuje role, obowiązki i procesy, mające na celu wspomaganie i propagowanie efektywnego wykorzystania BI w przedsiębiorstwie.

3.2.5. Analiza granulacyjna

Innym ważnym czynnikiem wpływającym na analizy BI jest granulacja (ziarnistość) informacji. Ziarna informacji stanowią elementy, które połączone są razem ze względu na ich: podobieństwo, nierozróżnialność, spójność, funkcjonalną bliskość czy bliskie położenie. Wskaźniki efektywności sprzedaży mogą dotyczyć np. określonego okresu (ziarnistość w czasie), terytorium (ziarnistość w przestrzeni), sprzedawców, asortymentu etc. Pojęcie granulacji ma oczywiście bardzo szeroki charakter i może być definiowane oraz interpretowane w rozmaity sposób: w sensie zbiorów rozmytych [Zadeh, 1979, 1997], zbiorów przybliżonych [Pawlak, 1982], systemów sąsiedzkich [Lin, 1997]. W bieżącym rozdziale będziemy je rozumieli w sensie wymiaru z analizy wielowymiarowej. Tak więc granulą będziemy określali ziarno informacji. Im mniejsze ziarno (mniejsza szerokość ziarna), tym bardziej precyzyjna jest informacja. Możemy więc podać wskaźnik sprzedaży jako wartość precyzyjną albo określić ją mniej precyzyjnie, jako satysfakcjonującą, niedostateczną etc. w zależności od sposobu definiowania ziarna. Za-

uważmy, że granule mogą być rozłączne lub mogą się pokrywać i że można tworzyć wiele perspektyw granulacji wskaźników efektywności, wykorzystując hierarchiczną strukturę granul.

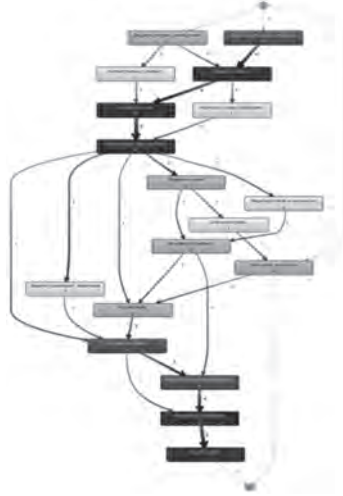
Systemy BI traktują ziarno informacji w kategoriach wartości liczbowej KPI. Nie istnieją jednak żadne przesłanki, które uniemożliwiałyby zastosowania ziaren w kategoriach np. wartości zbiorów przybliżonych.

3.2.6. Wizualizacja procesu

Kolejnym ważnym elementem analiz BI jest wizualizacja rzeczywistego przebiegu procesów biznesowych. Stan procesu można rozumieć w sensie wskaźników (temu zagadnieniu poświęcimy osobny rozdział „Wizualizacja i raportowanie”), ale także w sensie grafu procesu. Dane o zdarzeniach biznesowych, na podstawie których tworzony jest graf, zapisywane są wówczas na etapie wykonania procesu i stanowią chronologiczny zapis aktywności aplikacji/systemu IT wystarczający do rekonstrukcji, przeglądu oraz oceny sekwencji zdarzeń [Molski, 2007]. Obejmują one zazwyczaj takie informacje jak: zasoby (osoby lub urządzenia) wykonujące lub inicjujące krok procesu, znacznik czasu, dane związane z realizacją kroku procesu (np. wielkości zamówienia). Dane te wykorzystywane były do tej pory głównie przez audytorów systemów informatycznych, administratorów systemów i administratorów bezpieczeństwa informacji do zrozumienia przebiegu procesu, kontroli wewnętrznej oraz identyfikacji słabych punktów procesu. Obecnie mogą być traktowane jako wsparcie zarządzania procesami. Ziarnistość informacji można więc sprowadzić do pojedynczego zdarzenia procesu i ocenić KPI dotyczące np. zużycia energii elektrycznej, wody, gazu ziemnego oraz innych mediów w odniesieniu do wybranych zdarzeń procesu. Już sam model graficzny procesu daje decydentowi obraz stanu procesu, tak jak on wygląda w rzeczywistości. Na rysunkach 3.3 i 3.4 widzimy wizualizację procesu zamówienia sprzedaży. W procesie tym rejestrowane są zapotrzebowania klientów na nasze towary. Jeśli zamówienia są zatwierdzone, wówczas stanowią podstawę do utworzenia listy otwartych zamówień, na podstawie których możliwe jest wygenerowanie np. zlecenia produkcyjnego etc. Istnieje wiele możliwych realizacji procesu zamówienia sprzedaży. Graf, taki jak na rysunkach 3.3 i 3.4, pokazuje nam wszystkie możliwe realizacje procesu, ale daje też możliwość odfiltrowania realizacji nietypowych, identyfikacji wąskich gardeł procesu, identyfikacji zdarzeń istotnych z punktu widzenia przyjętych miar efektywności procesu etc. Daje też możliwość analizy częstości wykonywanych czynności procesu (rys. 3.3), np. zdarzeń niekorzystnych

w realizacji procesu lub analizie czasu realizacji poszczególnych czynności procesu oraz czasu oczekiwania na realizację kolejnej czynności (rys. 3.4). Większe wartości metryk reprezentowane są ciemniejszym kolorem.

Rys. 3.3. Częstość realizacji poszczególnych zdarzeń w procesie biznesowym



Źródło: opracowanie własne.

Rys. 3.4. Czas realizacji poszczególnych zdarzeń w procesie biznesowym



Źródło: opracowanie własne.

3.3. Wielowymiarowy model danych

Pamiętamy z rozdziału 3.1, że systemy BI charakteryzuje granulacja (ziarnistość) informacji. W roku 1993 Edgar F. Codd – ten sam, który w 1974 r. stworzył model relacyjny – zaproponował, na potrzeby przetwarzania informacji o różnej ziarnistości, rozszerzenie modelu relacyjnego o paradygmat modelu wielowymiarowego. W modelu wielowymiarowym informacje przechowywane są zatem – tak jak w modelu relacyjnym – w tabelach (tabelach faktów), które posiadają takie same własności jak te opisane w rozdziale 5. Dla tabel faktów przyjęto, że faktami (podstawowymi granułami informacji) nazywać się będzie atrybuty tabeli, natomiast miarami domeny atrybutów (faktów). Dodatkowo jednak w tabeli faktów pojawia się konieczność zdefiniowania stopnia ziarnistości faktów oraz hierarchii ziaren. Osiągnięto to poprzez dodanie do tabeli faktów kluczy obcych odwołujących się do kluczy głównych z tabel wymiarów. Wymiary są to zatem definicje ziarnistości informacji będące atrybutami w tabelach wymiarów, które ustalają kontekst analizy informacji. Każdy fakt można rozpatrywać w przestrzeni wielowymiarowej definiowanej przez wymiary. Dodatkowo jednak każdy wymiar może mieć swoją hierarchię.

Model wielowymiarowy wykorzystywany jest do budowy hurtowni danych. Sposób połączenia tabel faktów i tabel wymiarów w hurtowni danych ustalany jest na poziomie schematu hurtowni danych. Wyróżnia się trzy podstawowe schematy hurtowni danych:

1. Schemat gwiazdy;
2. Schemat płątka śniegu;
3. Schemat konstelacji.

W schemacie gwiazdy tabela faktów reprezentuje korpus gwiazdy, pojedyncze tabele wymiarów – ramiona gwiazdy. W schemacie płątka śniegu tabela faktów reprezentuje korpus płątka, wiele tabel wymiarów – ramiona płątka o określonej strukturze. Schemat płątka śniegu daje zatem projektantowi większą swobodę w definiowaniu hierarchii ziaren informacji. W schemacie konstelacji wiele tabel faktów współdzieli te same tabele wymiarów. Mamy zatem do czynienia z rozbudowanym korpusem gwiazdy lub korpusem płątka śniegu.

Przykładowo, jeśli w tabeli faktów przechowywane są informacje na temat sprzedaży netto (KPI), to mając zdefiniowaną hierarchię towaru w tabelach wymiarów, możliwe jest szybkie uzyskanie wartości KPI dla ustalonego poziomu hierarchii (określonych towarów lub grup towarów), stosując odpowiednią instrukcję języka SQL. Za każdym razem jednak

istnieje konieczność wykonania obliczeń na określonym poziomie zianowości informacji. Aby przyspieszyć uzyskanie odpowiedzi na zapytanie o określoną granulę informacji, można wykonać wstępne obliczenia i wprowadzić do hurtowni danych tzw. agregaty – zestawienia tworzone na określonym poziomie hierarchii granuli informacji. Dane zagregowane przechowywane są w tabelach wielowymiarowych (tzw. kostkach danych).

3.4. *Business Intelligence* na poszczególnych poziomach piramidy informacyjnej

3.4.1. Poziom danych

Dane, do których ma dostęp współczesna organizacja, to nie tylko tradycyjne relacyjne bazy danych, o których mowa w rozdziale Bazy danych, ale i nierelacyjne bazy danych (dla których używana jest nazwa NoSQL, co oznacza „no SQL”, „not have SQL”, „not relational SQL”, „non SQL” albo „not only SQL”, tzn. „nie SQL-owe” albo „nie tylko SQL-owe”). Pojęcie baza NoSQL obecnie utożsamiane jest z każdym innym rodzajem baz danych, poza bazą relacyjną. Zazwyczaj wyróżniane są cztery typy nierelacyjnych baz danych:

1. Bazy kolumnowe (przechowywane w całości w pamięci RAM, np. SAP HANA);
2. Bazy dokumentowe (np. IBM Domino, MongoDB);
3. Bazy oparte na kluczach i wartościach (bazy KVS – Key Value Storage, np. Oracle NoSQL Database);
4. Bazy grafowe (np. AllegroGraph);
5. Bazy oparte na kilku modelach jednocześnie (np. CortexDB).

Wykorzystywane są one jako np. wewnętrzne dane procesowe z sieci IIoT (*Industrial Internet of Things*) charakteryzujące stan i wydajność maszyn i urządzeń produkcyjnych, dane ze sklepów internetowych, dane z systemów CRM, strumienie danych przetwarzane w czasie rzeczywistym, dane tekstowe, dane z rozwiązań *Big Data* i in. Stanowią one źródło danych dla konsumentów produktów analitycznych/*Business Intelligence*, aplikacji analitycznych/biznesowych, narzędzi analitycznych i innych użytkowników⁸. Do integracji, modelowania, zarządzania,

⁸ Dresner Advisory Services, dresneradvisory.com.

przechowywania i uzyskiwania dostępu do tych danych służy analityczna infrastruktura danych (*Analytic Data Infrastructure*, ADI), która jest niezależna od narzędzi użytkownika; można ją nabyć osobno i uzyskać dostęp za pośrednictwem opublikowanych interfejsów API i DML⁹.

3.4.2. Poziom informacji

Na poziomie informacji wybierane i definiowane są – na podstawie zgromadzonych danych – wskaźniki efektywności oraz tworzona jest reprezentacja granularyczna wskaźników (grupowanie wskaźników w granule i ich hierarchie). Wymaga to semantycznej interpretacji umieszczenia wskaźników w jednej granuli i ustalenie relacji między granulami, stworzenia struktur granularycznych i perspektyw granulacji.

We wsparciu tego poziomu najważniejsze znaczenie mają modele wielowymiarowe, pozwalające na odzwierciedlenie zależności między granulami i narzędzia do budowy hurtowni danych i kostek analitycznych.

Hurtownia danych to zwykle dodatkowa baza danych SQL-owa w przedsiębiorstwie zbudowana ze znormalizowanej tabeli faktów i zde-normalizowanych tabel wymiarów, zawierająca zintegrowane dane historyczne pochodzące z rozproszonych systemów bazodanowych, standardowo wykorzystująca schemat gwiazdy (lub niekiedy schemat płątka śniegu), wykorzystywana do analiz granulacyjnych.

Tabela faktów jest kolekcją miar liczbowych (wskaźników efektywności) charakteryzujących proces biznesowy oraz kluczy obcych do tabel wymiarów. Tabele wymiarów są kolekcją struktur hierarchicznych charakteryzujących wymiary (perspektywy), według których można przeglądać wartości miar, co umożliwia tworzenie reprezentacji granularycznej wskaźników efektywności. W hurtowni danych zazwyczaj występuje wiele tabel faktów i wiele tabel wymiarów, a każda tabela faktów ma zdefiniowanych wiele miar.

Członkowie wymiaru definiują perspektywy, w jakich będzie można oglądać i analizować dane z tabeli faktów (np. w perspektywie koloru, rozmiaru, wagi etc.).

Klucze obce świadczą o tym, że wymiar ma rozbudowaną formę normalną (z dodatkowymi tabelami charakteryzującymi wymiar, np. występuje tabela *DimSubcategory*, o czym świadczy występowanie klucza *SubcategoryKey*).

9 *Ibidem*.

Ponieważ hurtownia danych jest bazą relacyjną, informacje z hurtowni danych mogą być pobierane poprzez skrypty SQL-owe. Do analizy danych pochodzących z hurtowni danych często wykorzystywane są kostki danych/kostki OLAP (*On-Line Analytical Processing*) i rozwiązania analityczne zbudowane wokół kostki danych. Kostka OLAP zawiera dane wielowymiarowe i łączy wartości miar i wymiarów w jeden model pojęciowy.

Hurtownie danych oraz rozwiązania analityczne zbudowane wokół hurtowni danych (systemy klasy OLAP, systemy eksploracji danych, o których będzie mowa w dalszej części rozdziału i kokpity menedżerskie) składają się na projekty *Business Intelligence*, które mogą stanowić integralną część systemów ERP lub mogą być projektami klasy Enterprise. Ze względu na wymagane zasoby finansowe i zasoby ICT w projekcie decyzja o uruchomieniu projektu *Business Intelligence* ma dla przedsiębiorstwa charakter strategiczny i wiąże się z wprowadzeniem zarządzania przez dane. Łączy się to z problematyką zarządzania projektami i integracją projektu z celami strategicznymi przedsiębiorstwa.

Na poziomie informacji realizowane są analizy i przeszukiwanie struktur granulacyjnych. Wymaga to określenia narzędzi do analiz wielowymiarowych. Stosowane tu narzędzia to m.in. MDX i tabela przestawna oraz eksploracja danych. Istotne jest tu określenie stanu przedsiębiorstwa poprzez ustalenie stosownych wartości KPI oraz przeprowadzenie analizy porównawczej.

3.4.3. Poziom wiedzy

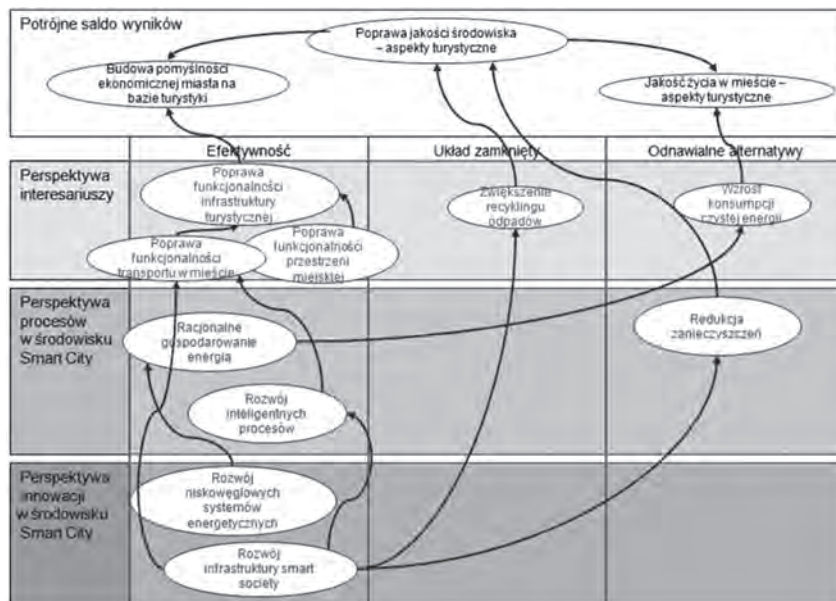
Poziom wiedzy to połączenie celów strategicznych organizacji z informacją dostępną na poziomie informacyjnym. Kaplan i Norton w roku 1996 wprowadzili zrównoważoną kartę wyników opartą na systemie pomiaru efektywności działań operacyjnych w odniesieniu do strategii, który umożliwia dekompozycję strategii na procesy/projekty. System pomiaru zrównoważonej karty wyników grupuje wskaźniki efektywności w cztery kategorie: finansową, klienta, procesów wewnętrznych, infrastruktury i rozwoju. Koncepcja zrównoważonej karty wyników zasadza się na idei wymuszenia mierzenia efektów działań strategicznych i została sformułowana w wyniku projektu badawczego „Mierzenie efektywności w organizacjach przyszłości” [Kaplan, 1992].

Punktem wyjścia do opracowania zrównoważonej karty wyników jest zazwyczaj ustalenie łańcucha zależności przyczynowo-skutkowych pomiędzy celami strategicznymi w formie mapy strategii. Na rysunku 3.5

prezentujemy zmodyfikowaną na potrzeby książki (w stosunku do propozycji Kaplana i Nortona) mapę strategii, tak aby uwzględniała ona archetypy innowacyjnych modeli biznesu na rzecz inteligentnego i zrównoważonego miasta: maksymalną efektywność materiałową i energetyczną, gospodarkę zasobami w układzie zamkniętym, zastępowanie odnawialnymi alternatywami, dostarczanie funkcjonalności zamiast własności, przyjmowanie funkcji włodarza, spowolnienie konsumpcji, zaangażowanie interesariuszy do współtworzenia wartości, promowanie innowacyjnych rozwiązań i praktyk poprzez przedsiębiorczość społeczną.

Rysunek 3.5 pokazuje w układzie mapy strategii, które cele ekonomiczne, społeczne i środowiskowe projektów są powiązane ze sobą w ramach wspólnego celu strategicznego, jaki ma zostać osiągnięty dzięki ich realizacji.

Rys. 3.5. Mapa strategii – ilustracyjny przykład



Źródło: [Gontar, 2016].

Dla każdego z ww. celów ustalane są mierniki osiągnięcia celów strategicznych (KPI), docelowe wartości mierników, działania służące osiągnięciu celów strategicznych. Przełożenie celów strategicznych na konkretne działania przedsiębiorstwa wymaga zdefiniowania składu portfela projektów. Tu już wchodzimy na poziom mądrości strategicznej organizacji.

3.4.4. Poziom mądrości strategicznej

Poziom mądrości strategicznej implementacji polega na wykonaniu akcji związanych z interpretacją otrzymanej wiedzy i wprowadzaniem w życie rozwiązania problemu dopasowania działań przedsiębiorstwa do pożądaných celów strategicznych. Wybór działań do realizacji wymaga zastosowania modeli decyzyjnych (zwykle wielokryterialnych). Należy zwrócić uwagę na to, że przyjęta metoda oceny i wartościowania projektów może zakładać równorzędność celów ekonomicznych, społecznych i środowiskowych, ale może też dać sposobność ustalenia priorytetów. Na portfel projektów składają się projekty pogrupowane ze sobą w taki sposób, by uzyskać portfele, które realizują jak największą liczbę równorzędnych celów. Z jednej strony zatem chcielibyśmy otrzymać taki skład portfela projektów, który odnosi się do jak największej liczby celów, ale z drugiej strony chcielibyśmy mieć w portfelu takie projekty, które są najbardziej efektywne we wspomaganii celów.

Wartościowanie projektów wspomagających strategię przedsiębiorstwa i analiza ich wartości ze względu na efekty ekonomiczne, społeczne i środowiskowe przyjmuje formę analizy obwiedni danych (DEA), analizy hierarchicznej problemu (AHP), analizy z logiką rozmytą, analizy z wykorzystaniem algorytmów ewolucyjnych etc.

3.5. Charakterystyka systemów *Business Intelligence*

3.5.1. Pojęcie systemu *Business Intelligence*

Systemy *Business Intelligence* są zbiorem ludzi, strategii i procesów, technologii i narzędzi, w tym hurtowni danych oraz hurtowni tematycznych, które integrują dane i przekształcają je w użyteczne informacje, które pomagają organizacji zrozumieć przeszłość i kształtować jej przyszłe wyniki [CAM-I, 1992].

3.5.2. Cechy systemów BI

Systemy BI nazywane są często współczesną wersją systemów wspomagania decyzji, o których piszemy w drugiej części podręcznika. Wykorzystamy zatem opis cech systemów DSS, dostosowując go do bieżących możliwości systemów BI. Analizując zadania BI, możemy powiedzieć, że są to systemy informatyczne, które wspomagają decydentów w obliczaniu i analizie wskaźników efektywnościowych z wykorzystaniem modeli prezentacyjnych i analitycznych, bazując na danych pochodzących z hurtowni danych/hurtowni tematycznych oraz w podejmowaniu decyzji, które mają umożliwiać osiągnięcie zakładanych celów. Specyficzne cechy systemów BI to:

1. Zakres zastosowań. BI dostarcza informacji wspomagających kompleksowo proces zarządzania daną dziedziną (np. hurtownie tematyczne i systemy CRM). BI wspierać może także całą organizację (hurtownie danych i kompleksowe systemy BI);
2. Konieczność obliczania KPI z uwzględnieniem ich ziarnistości, wykonywania na ziarnach KPI o różnej szerokości skomplikowanych analiz i porównań, przy wykorzystaniu zaawansowanych pakietów oprogramowania. Przede wszystkim wymagane jest zastosowanie analizy porównawczej, co nakłada na organizację wdrażającą systemy BI obowiązek realizacji zadań związanych z wywiadem gospodarczym. Dalej analiza stanu przedsiębiorstwa wymaga prezentacji i analizy KPI na różnych poziomach ziarnistości informacji, co nakłada na systemy BI obowiązek zapewnienia możliwości drążenia informacji w górę i w dół w celu przejrzenia wyższych i niższych poziomów szczegółowości informacji, ale także dopuszcza możliwość wizualizacji procesów biznesowych w postaci grafów, czyli zbiorów węzłów i łączących je krawędzi. Wreszcie diagnoza stanu przedsiębiorstwa ma pomóc w wypracowaniu decyzji, które sprzęgnięte są z celami przedsiębiorstwa. W związku z tym Systemy BI tworzone są często w specjalistycznych środowiskach programistycznych dostarczających odpowiednich procedur obliczeniowych (takich jak np. SAS, Matlab czy nawet w prostszych przypadkach MS Excel). BI może również – tak jak w systemach DSS – stanowić platformę integrującą dla samodzielnych zewnętrznych programów analitycznych;
3. Możliwość przetwarzania dużych ilości danych pochodzących z różnych źródeł i tworzenie na ich podstawie hurtowni danych i hurtowni tematycznych, które zawierają dane potrzebne do obliczeń wskaźników efektywnościowych KPI, albo wręcz przechowują wartości

KPI, i które są podstawą do zaawansowanych analiz i wizualizacji. BI wyspecjalizowane w analizie dużych zbiorów danych (*Big Data*) oraz analizujące zależności między danymi przy pomocy metod granularnych (*granular analysis*) stanowią jedną z najbardziej obiecujących dziedzin informatyki;

4. BI opierają się zatem przede wszystkim na interakcyjnych raportach na ekranie komputera, pozwalających na nawiązanie dialogu użytkownika z systemem i współpracę z nim w trybie on-line w kokpitach menedżerskich (zarówno statycznych, jak i dynamicznych);
5. Analiza porównawcza, czyli *benchmarking* (zwany także „benchmarkingiem najlepszych praktyk” lub „benchmarkingiem procesów”). Systemy BI pozwalają porównywać procesy biznesowe i wskaźniki wydajności z najlepszymi w branży lub najlepszymi praktykami z innych branż. Zazwyczaj przedmiotem porównań są: jakość, czas i koszt. Analiza porównawcza może być wydarzeniem jednorazowym, ale często jest traktowana jako ciągły proces, w którym organizacje nieustannie dążą do poprawy swoich praktyk.

3.5.3. Struktura systemów BI

Systemy BI tworzone są z myślą o wspomaganiu kompleksowo procesu zarządzania daną dziedziną. W związku z tym trudno jest mówić o jakimś uniwersalnym jednolitym wzorcu ich budowy. BI mają silnie zróżnicowaną strukturę wewnętrzną, w zależności od dziedziny, dla której zostały one stworzone, oraz preferencji użytkownika odnośnie do działania systemu. Tym niemniej możemy wyróżnić pewien ramowy schemat ich budowy, obejmujący kilka standardowych podsystemów:

1. System pomiarów wskaźników efektywności procesów biznesowych w organizacji, w tym *benchmarking*, który pozwala na ustalenie poziomu efektywności procesów w zależności od wartości wskaźników względem modelowych organizacji, np. najsilniejszych konkurentów. Wskaźniki efektywności weryfikują osiągnięcie celów strategicznych i często odwołują się do zrównoważonej karty wyników. Do integracji, modelowania, zarządzania, przechowywania i uzyskiwania dostępu do danych wykorzystywanych do obliczania wskaźników efektywności służy analityczna infrastruktura danych, która jest niezależna od narzędzi użytkownika, można ją nabyć osobno i uzyskać dostęp za pośrednictwem opublikowanych interfejsów API i DML. Należy pamiętać o tym, że mierniki oceny efektywności procesów biznesowych ewoluują wraz z rozwojem

celów strategicznych organizacji, ale także wraz z rozwojem teorii zarządzania strategicznego. Dane do obliczania wskaźników efektywności przechowywane są w bazach danych analitycznych (jedna hurtownia danych lub wiele hurtowni tematycznych) i zawierają aktualne dane dotyczące działalności organizacji i jej otoczenia. Z tego powodu możemy czasami mówić o bazie danych wewnętrznych i zewnętrznych organizacji. Dane wewnętrzne pochodzą przede wszystkim z baz danych transakcyjnych oraz innych systemów informatycznych. Dane zewnętrzne pochodzą zwykle od otoczenia gospodarczo-politycznego organizacji;

2. Baza porównawcza (referencyjna) stanowiąca punkt odniesienia do oceny stanu przedsiębiorstwa. Zawiera dane historyczne lub aktualne dane konkurencji zebrane metodami wywiadu gospodarczego;
3. Podsystem analityczny pozwalający na podstawie informacji z baz danych i baz porównawczych uzyskać wiedzę o stanie przedsiębiorstwa. Wykorzystywane są tu zazwyczaj analizy OLAP (wielowymiarowe) oraz eksploracja danych (realizowana także w wielu wymiarach). Konieczne jest także zastosowanie systemu informacji strategicznej, który pozwoliłby pokazać łańcuch przyczynowo-skutkowy celów strategicznych oraz przypisałby poszczególnym celom określone wskaźniki KPI;
4. Podsystem symulacyjno-decyzyjny wspomagający zarządzanie wiedzą, który na podstawie żądań użytkownika oraz istniejących danych dokonuje wyboru kombinacji modeli niezbędnych do rozwiązania zadania, wyboru danych wejściowych dla tych modeli oraz wykonuje z ich pomocą niezbędne obliczenia. Zauważmy, że przepływy informacyjne tego podsystemu z bazą danych mają charakter dwukierunkowy. Może on nie tylko pobierać dane wejściowe, ale również zapisywać w bazie danych informacje będące wynikiem działania BI. Podobnie w przypadku bazy modeli podsystem symulacyjno-decyzyjny może również modyfikować modele pod kątem konkretnego problemu oraz istniejących danych (np. poprzez reestymację ich parametrów), zapisując zmiany w bazie modeli;
5. Platforma komunikacyjna ma zapewnić wysoki komfort obsługi i możliwość wypracowania wspólnych decyzji.

3.6. Języki programowania systemów BI

Praca z systemami BI wymaga od zaawansowanego użytkownika znajomości języków programowania, np. SQL, który został opisany w rozdziale 2. Język SQL nie wspiera jednak w wystarczającym stopniu analityki danych, co jest przesłanką działań podejmowanych przez dostawców rozwiązań bazodanowych, narzędzi developerskich i aplikacji biznesowych w zakresie rozszerzeń języka SQL o funkcje analityczne. Warto w tym miejscu wspomnieć o trzech próbach Microsoftu: językach DAX, DMX i MDX. Ze względu na ograniczoną objętość książki w dalszej części rozdziału opisany zostanie pokrótce język DMX. Należy też pamiętać o tym, że języki takie jak DAX, DMX i MDX oferują pewien standard za cenę małej elastyczności, a tym samym ograniczają obszar zastosowań do tego, który jest zdefiniowany w standardzie. W językach tych można więc przetwarzać ziarna informacji tylko w takiej formie, w jakiej dopuszcza to standard. Nie można zatem – korzystając z tych języków – wykorzystać np. zbiorów rozmytych, zbiorów przybliżonych etc.

Język DMX (*Data Mining eXtension*) został opublikowany w roku 2000 w firmie Microsoft jako element standardu OLE DB for Data Mining i został zaimplementowany po raz pierwszy w komercyjnym produkcie w systemie Microsoft SQL Server 2000. Obecnie stosowany jest w systemach Microsoft SQL Server Analysis i Reporting Services. Przyjrzymy się pobieżnie temu językowi ze względu na popularność tego rozwiązania w Polsce i pozycję Microsoftu na rynku systemów *Business Intelligence*.

Standard OLE DB for DM wywodzi się z dwóch technologii bazodanowych: OLE DB oraz języka SQL. DMX nie jest – w przeciwieństwie do języka SQL – standardem powszechnie uznawanym i stosowanym. Używany jest głównie przez projektantów i programistów baz danych, a także przez nieprofesjonalnych, ale zaawansowanych użytkowników.

Język DMX jest narzędziem całkowicie wystarczającym do pracy bezpośredniej (konwersacyjnej) z tabelarycznymi zbiorami danych w zakresie analiz eksploracyjnych. Poniżej przedstawiono przykładową, prostą instrukcję tworzącą strukturę Pros_Cons z dwoma wbudowanymi tabelami pomocniczymi Pros i Cons. Pros oznacza korzyści (pozytywne konsekwencje – tab. 3.4), a Cons (tab. 3.5) bariery transformacji do Przemysłu 4.0.

Tabela 3.4. Transformacja do Przemysłu 4.0 – pozytywne konsekwencje

1. Niższe koszty operacyjne (związane z produkcją)
2. Lepsze dostosowanie do oczekiwań klientów i partnerów biznesowych
3. Lepsze wykorzystanie danych do zarządzania wiedzą
4. Rozwój nowych produktów i tworzenie innowacji
5. Opracowanie nowych modeli biznesowych
6. Lepsza komunikacja z otoczeniem
7. Zwiększenie wydajności pracy
8. Skrócenie czasu wprowadzenia produktu na rynek
9. Większa elastyczność oferowanych produktów i usług
10. Pozwala skorzystać z aplikacji do tej pory niedostępnych z uwagi na wysoki koszt
11. Efektywniejsza alokacja zasobów
12. Usprawnienie przygotowania produkcji
13. Usprawnienie procesów produkcji
14. Usprawnienie procesów zarządzania jakością
15. Szybki dostęp do wiedzy
16. Dostęp do szczegółowych danych związanych z procesami produkcji
17. Szybsze i wydajniejsze przetwarzanie danych
18. Swobodny dostęp do danych i aplikacji z dowolnego miejsca
19. Zastąpienie kilku aplikacji jednym środowiskiem
20. Zagwarantowanie bezpieczeństwa (danych i systemów)
21. Zapewnienie lepszej ciągłości działania systemów informatycznych
22. Szybki dostęp do nowo wprowadzanych technologii ICT
23. Efektywniejsza aktualizacja aplikacji
24. Redukcja problemów z utrzymaniem infrastruktury
25. Większa możliwość dopasowania funkcjonalności systemów do procesów biznesowych

Źródło: opracowanie własne.

Tabela 3.5. Transformacja do Przemysłu 4.0 – bariery

1. Kwestie ochrony danych osobowych
2. Aspekty prawne
3. Ograniczone zaufanie do nowych technologii
4. Obawy o bezpieczeństwo danych i usług
5. Obawa przed działaniami dostawców usług chmurowych
6. Koszty wdrożenia
7. Trudności z integracją rozwiązań
8. Obawy o jakość usług świadczonych przez dostawców chmury
9. Trudny dostęp wskutek słabego łącza
10. Brak wiedzy i kompetencji pracowników firmy
11. Nieznany wpływ na zarządzanie firmą
12. Złożoność umów SLA (umowa o świadczenie poziomu usługi)

Źródło: opracowanie własne.

Chcemy ustalić, które bariery/korzyści występują wspólnie na drodze do Przemysłu 4.0.

```
CREATE MINING STRUCTURE Pros_Cons
(
  [CaseID] long key,
  [Pros] table
  (
    [Pro] text key
  ),
  [Cons] table
  (
    [Con] text key
  )
);
```

W dalszej części rozdziału zapoznamy się ze składnią zapytań w języku skryptowym DMX (CREATE, ALTER, EXPORT, IMPORT, DROP, INSERT INTO, SELECT, DELETE) oraz postaramy się zrozumieć pojęcia definiowane w skrypcie DMX: struktura eksploracji (*Mining Structure*), model eksploracji (*Data Mining Model*), przypadki (*Cases*), przypadki zagnieżdżone (*Nested Tables*), typy danych kolumny (*Column Data Types*), typy zawartości kolumny (*Content Types*), użycie kolumn (*Column Usage*), rozkłady kolumn, dyskretyzacja danych. Skrypt DMX-owy uruchamia następujący proces eksploracji danych:

1. Zdefiniowanie struktury eksploracji (CREATE MINING STRUCTURE);
2. Dołączenie danych (INSERT INTO);
3. Dołączenie modelu (ADD MINING MODEL);
4. Uczenie modelu z wykorzystaniem dołączonych danych (INSERT INTO);
5. Predykcja (SELECT).

Eksploracja danych w języku DMX realizowana jest na strukturach danych określanych jako przypadki (*Cases*). Elementami struktury są kolumny, które mogą mieć strukturę prostą lub złożoną. Jeśli kolumna przypadku ma strukturę prostą, to elementy kolumny mają zdefiniowany typ danych oraz typ zawartości. W przeciwnym przypadku kolumna ma strukturę złożoną i elementami kolumny są tabele zagnieżdżone (*Nested*). Oczywiście elementy tabeli zagnieżdżonej także mają zdefiniowany typ danych oraz typ zawartości.

Aby zrealizować proces eksploracji danych w języku DMX, musimy więc mieć zdefiniowaną strukturę eksploracji.

Struktura eksploracji:

1. Opisuje kolumny przypadków (*Cases* oraz *Nested*) poprzez zdefiniowanie:
 - a) typów danych (*Column Data Type*),
 - b) typów zawartości (*Content Types*),
 - c) rozkładu,
 - d) metod dyskretyzacji,
 - e) powiązań z innymi kolumnami,
 - f) flag;
2. Definiuje modele eksploracji;
3. Przechowuje dane treningowe (*Cases*, *Nested*);
4. Przechowuje wydzielone dane testowe (*Holdout*);
5. Przechowuje rezultaty eksploracji.

W języku DMX strukturę eksploracji definiujemy poprzez wywołanie zapytania DMX-owego (`CREATE MINING STRUCTURE`). Ogólnie, składnia zapytania (`CREATE MINING STRUCTURE`) przedstawia się następująco:

```
CREATE MINING STRUCTURE [<nazwa struktury eksploracji>]
(
  <kolumna klucza>,
  <kolumny struktury eksploracji>,
  <tabela zagnieżdżona>
  (<kolumna klucza tabeli zagnieżdżonej>,
  <kolumny struktury eksploracji tabeli zagnieżdżonej> )
)
```

Omówimy pokrótce poszczególne elementy.

`Pros_Cons` to nazwa struktury eksploracji. Jest to nazwa własna, którą nadajemy strukturze eksploracji. Musi być ona zdefiniowana zgodnie z zasadami określonymi w MSDN-ie w dziale Identifiers (DMX). Zauważmy, że nazwa może zawierać spacje i/lub słowa zastrzeżone (np. `model`). W takim przypadku jednak nazwę należy ująć w nawiasy kwadratowe (np. [`Pros & Cons`]).

[`CaseID`] to kolumna klucza – czyli kolumna jednoznacznie identyfikująca encję w przypadkach (*Cases*). Przyjmujemy następującą zasadę: dane z przypadków (*Cases*) pochodzą z jednej tabeli, natomiast dane ze struktur zagnieżdżonych (*Nested*) mogą pochodzić z innej/innych tabel. Zauważmy na koniec, że kolumna klucza w strukturze eksploracji jest niezależna od klucza w tabelach źródłowych. Oznacza to, że w tabeli może wystąpić więcej encji z tą wartością klucza, natomiast w przypadkach (*Cases*) te przypadki redukują się do jednego wystąpienia.

W naszym przykładzie nie było – oprócz kolumny klucza – innych kolumn struktury eksploracji. Kolumny eksploracji mają zdefiniowaną nazwę, typ danych oraz typ zawartości. Oczywiście, typ danych określa zakres danych, który może być przypisany do kolumny struktury eksploracji, oraz operacje, które na tych danych można zrealizować. Typ danych w strukturze eksploracji definiujemy tak, aby zapewnić zgodność typu kolumny eksploracji danych z typem danych kolumny tabeli/widoku bazy danych w tabelach źródłowych.

Podstawowe typy danych dostępne w języku DMX to:

1. Text
2. Long
3. Boolean
4. Double
5. Date

Typ zawartości zdefiniowany jest na potrzeby algorytmów eksploracji danych. Podstawowe typy zawartości dostępne w języku DMX to:

1. DISCRETE
2. CONTINUOUS
3. DISCRETIZED
4. KEY
5. KEY SEQUENCE
6. KEY TIME
7. ORDERED
8. CYCLICAL

Typ zawartości dla kolumny klucza to oczywiście KEY. Do pozostałych typów zawartości wrócimy, omawiając następne przykłady.

Pros oraz Cons to dane zagnieżdżone (*Nested*). Encja w danych do analizy zawiera zatem pojedyncze elementy kolumny klucza oraz struktury eksploracji z przypadków (*Cases*) oraz tabele ze struktur zagnieżdżonych (*Nested*). W naszym przypadku wyglądałoby to w następujący sposób:

| [CaseID] | [Pros] |
|----------|---|
| 1 | Niższe koszty operacyjne (związane z produkcją) Szybki dostęp do wiedzy Dostęp do szczegółowych danych związanych z procesami produkcji |
| ... | |
| 2 | Lepsza komunikacja z otoczeniem Zwiększenie wydajności pracy |
| ... | |

Oczywiście pierwsza kolumna zawiera [CaseID], a druga tabelkę [Pros].

Jak należy to zatem czytać? Widzimy, że do analizy mamy dane o korzyściach transformacji do Przemysłu 4.0 w kolejnych zakładach. Chcemy zatem dowiedzieć się, jakie korzyści są zazwyczaj wspólnie dzielone w przedsiębiorstwie. Pro to kolumna klucza tabeli zagnieżdżonej. Zauważmy, że podobnie do przypadków (*Cases*) w tabeli zagnieżdżonej (*Nested*) może dojść do sytuacji, że w tabeli skąd *Nested* czerpie dane – może wystąpić więcej encji z tą wartością klucza, natomiast w tabeli zagnieżdżonej (*Nested*) te przypadki redukują się do jednego wystąpienia.

INSERT INTO

INSERT INTO zostało wykorzystane do dołączenia danych. Zauważmy, że dane są dołączone do struktury eksploracji danych [Pros_Cons], a nie do modelu. Tym samym, jeśli w jednej strukturze zdefiniujemy więcej modeli, będą one korzystać z jednego zestawu danych.

Zauważmy też, że w dalszej części książki wywołaliśmy jeszcze raz INSERT INTO – tym razem na rzecz modelu Rekomendacje. W tej sytuacji, a także wówczas, gdy wywołujemy polecenie INSERT INTO na rzecz struktury, do której dołączyliśmy już model, po dołączeniu danych do struktury wywoływany jest proces dopasowania modelu do danych (inaczej – proces uczenia modelu).

INSERT INTO Rekomendacje

Ogólnie, składnia polecenia INSERT INTO jest następująca:

```
INSERT INTO MINING STRUCTURE [<nazwa struktury eksploracji>]
(
  <kolumny struktury eksploracji>
  [<tabela zagnieżdżona>]
  ( SKIP, <pominięte kolumny> )
)
SHAPE {
  OPENQUERY(['<źródło danych>'],'<zapytanie SELECT>') }
APPEND
(
  {OPENQUERY(['<źródło danych może_byc_inne>'],'<zapytanie
SELECT dla tabeli zagnieżdżonej>')}
)
RELATE [<klucz case>] TO [<klucz obcy>]
) AS [<tabela zagnieżdżona>]
-
```

Widzimy więc, że w poleceniu INSERT INTO należy zidentyfikować strukturę eksploracji (lub model eksploracji), podać listę kolumn

zdefiniowanych w strukturze eksploracji (uwzględniając tabele zagnieżdżone) i przypisać do nich dane treningowe (SELECT DISTINCT [CaseID] AS [CaseID]') oraz zdefiniować same dane treningowe. Jeśli wykorzystujemy tabelę zagnieżdżoną, wówczas w trakcie definiowania danych treningowych używamy klauzuli SHAPE. Wykorzystanie klauzuli SHAPE możemy przyrównać do ustanowienia relacji pomiędzy tabelami CASE i NESTED. Fizyczne relacje nie muszą jednak istnieć między tabelami.

Zauważmy też, że w drugim przykładzie użyliśmy klauzuli SKIP, aby zignorować kolumny z tabeli, których to kolumn nie zdefiniowaliśmy w strukturze eksploracji.

Powiedzmy jeszcze parę słów na temat klauzuli SHAPE. Ogólna składnia jest następująca:

```
SHAPE {<zapytanie do tabeli nadrzędnej >}
APPEND ({ <zapytanie do tabeli podrzędnej > }
  RELATE <kolumna tabeli nadrzędnej> TO <kolumna tabeli podrzędnej>)
  S <nazwa tabeli>
[
  ({ <zapytanie do tabeli podrzędnej > }
  RELATE <kolumna tabeli nadrzędnej > TO <kolumna tabeli podrzędnej>)
  AS <nazwa tabeli>
...
]
```

ALTER MINING STRUCTURE

Polecenie ALTER MINING STRUCTURE wykorzystywane jest do dołączenia modelu eksploracji danych do struktury.

```
ALTER MINING STRUCTURE Pros_Cons
ADD MINING MODEL Rekomendacje
(
  [CaseID],
  [Pros] PREDICT
  (
    Pro
  )
)USING Microsoft_Association_Rules
(MINIMUM_SUPPORT=10,MINIMUM_PROBABILITY=0.4)
```


Zauważmy, że można wyróżnić w tym poleceniu kilka wyraźnie zarysowanych części:

1. Odwołanie do struktury eksploracji Pros_Cons;
2. Ustalenie nazwy modelu eksploracji rekomendacje;
3. Odwołanie do kolumn z typem zawartości key [CaseID], Pro;
4. Odwołanie do kolumn z atrybutem input i predict Pro;
5. Odwołanie do tabeli zagnieżdżonej Pros;
6. Zdefiniowanie algorytmu eksploracji danych i jego parametrów.

Zauważmy, że w naszym przykładzie chcemy podać na wejściu listę korzyści, oczekując, że na wyjściu uzyskamy inną listę korzyści (zawierającą korzyści deklarowane w naszych zakładach wspólnie z wejściową listą korzyści). Ponieważ lista może zawierać więcej niż jedną korzyść, atrybut PREDICT przypisujemy całej tabeli Pros.

Uwaga: w naszym przykładzie tabela Pros ma atrybut zarówno Input, jak i Predict. Domyślnym atrybutem jest Input. Atrybuty Predict czy Predict Only nadajemy jawnie.

Oto składnia polecenia ALTER MINING STRUCTURE:

```
ALTER MINING STRUCTURE [<nazwa struktury eksploracji>]
ADD MINING MODEL [<nazwa modelu eksploracji>]
(
  [<kolumna klucza>],
  <kolumny modelu eksploracji>,
  <tabela zagnieżdżona>
  ( [<kolumna klucza tabeli zagnieżdżonej>],
    <kolumny modelu eksploracji tabeli zagnieżdżonej> )
) USING <algorytm>( <parametry algorytmu> )
```

Kolumnami modelu eksploracji mogą być tylko kolumny zdefiniowane w strukturze eksploracji. Oczywiście w kodzie DMX-owym możemy utworzyć model eksploracji bez wcześniejszego utworzenia struktury eksploracji (CREATE MINING MODEL). W tym przypadku jednak w tle tworzona jest struktura eksploracji z tymi samymi kolumnami, jakie zdefiniowaliśmy w modelu.

Zauważmy też, że pierwszą kolumną jest kolumna klucza. Dotyczy to też tabeli zagnieżdżonej. Pozostałe kolumny to kolumny modelu eksploracji.

Dodając model eksploracji, musimy odwołać się do predefiniowanej nazwy algorytmu USING Microsoft_Association_Rules. Każdy algorytm ma parametry o ustalonych wartościach domyślnych. Jeśli chcemy je zmienić, musimy jawnie to zapisać:

```
MINIMUM_SUPPORT=10,MINIMUM_PROBABILITY=0.4).
```

Ponowne wywołanie polecenia INSERT INTO utworzy nam instancję modelu dla danych treningowych.

INSERT INTO Rekomendacje

Jesteśmy więc gotowi do realizacji analizy predykcyjnej – czyli do głównego celu modelowania. Zbudowaliśmy bowiem model, aby odpowiedzieć sobie na pytanie, co nasze przedsiębiorstwo uważa za korzyść z transformacji do Przemysłu 4.0, jeśli już wiadomo, że mamy „Redukcję problemów z utrzymaniem infrastruktury” oraz otrzymaliśmy możliwość „Opracowania nowych modeli biznesowych”. Musimy zatem w analizie predykcyjnej dać rekomendację naszym menedżerom, czego powinni wymagać od naszego działu IT jako możliwe dodatkowe korzyści. Musimy zatem wywołać polecenie SELECT.

```
SELECT FLATTENED Predict(Pros, 5) FROM Rekomendacje  
NATURAL PREDICTION JOIN
```

```
(  
  SELECT (  
    SELECT ‚Redukcja problemów z utrzymaniem infrastruktury’ AS Pro UNION  
    SELECT ‚Opracowania nowych modeli biznesowych’ AS Pro  
  ) AS Pros  
) AS T
```

Jak widzimy, w poleceniu SELECT można wyróżnić kilka wyraźnych części:

1. Funkcję predykcyjną PREDICT (ogólnie jest to lista kolumn modelu eksploracji i funkcji predykcyjnych, a także kolumn ze zbioru wejściowego zawierającego dane do predykcji – będziemy to ćwiczyć przez cały semestr, zatem na pewno odwołamy się do każdej możliwości);
2. Zapytanie do danych predykcyjnych, czyli do danych, na podstawie których zrealizujemy analizę predykcyjną: NATURAL PREDICTION JOIN... Mamy tu dwie możliwości: albo wprowadzimy dane w zapytaniu, tak jak to widzimy w naszym przypadku, albo sięgniemy do danych poprzez natywną bazę danych, natywną kostkę OLAP-ową lub plik zewnętrzny. W tym drugim przypadku mamy kilka narzędzi do wyboru. Wymieńmy OPENQUERY, OPENROWSET i MDX;
3. Mapowanie pomiędzy kolumnami modelu eksploracji i danymi predykcyjnymi. W przypadku gdy nazwy są identyczne, możemy użyć klauzuli NATURAL i nie realizować mapowania.

Składnia polecenia SELECT (w wersji wywołanej w naszym przykładzie) jest następująca:

```
SELECT <lista kolumn modelu eksploracji, funkcji predykcyjnych
i danych predykcyjnych>
FROM [<model eksploracji >] NATURAL PREDICTION JOIN
(SELECT ,<wartość>' AS [<kolumna modelu eksploracji>],
(SELECT ,wartość' AS [<kolumna tabeli zagnieżdżonej modelu
eksploracji >] UNION
SELECT ,wartość' AS [<kolumna tabeli zagnieżdżonej modelu
eksploracji >] ...)
AS [<tabela zagnieżdżona>])
AS [<alias danych wejściowych>]
```

Dodajmy na zakończenie, że możemy w analizie predykcyjnej oszacować prawdopodobieństwa, że przedsiębiorstwo rzeczywiście odczuje wskazane korzyści. W tym celu wywołamy trochę inaczej funkcję PREDICT.

```
PREDICT(Pros, INCLUDE_STATISTICS, 5)
```

W efekcie, otrzymamy dodatkowo informacje dotyczące: wsparcia, prawdopodobieństwa i skorygowanego prawdopodobieństwa.

3.7. Podsumowanie

Korzyści z wykorzystania systemu BI leżą głównie w zakresie rozwoju cyfrowego DNA przedsiębiorstwa. Aby zrozumieć cyfrowe DNA przedsiębiorstwa, posłużmy się analogią z cyfrowym DNA człowieka odnoszącym się do techniki urządzeń ubieralnych. Jak podaje Ryszard Tadeusiewicz¹⁰, badania przeprowadzone przez *Pew Research Center Internet & American Life Project* wykazują, że 69% dorosłych Amerykanów śledzi jakiś wskaźnik swojego zdrowia lub zdrowia bliskiej osoby¹¹, np. wagę, dietę, ciśnienie krwi, poziom cukru we krwi, bóle głowy lub sen, za pomocą IoT. Analogicznie w inteligentnym i zrównoważonym przedsiębiorstwie istnieje pokusa nieustannego śledzenia wskaźników odnoszących się do „procesów życiowych” przedsiębiorstwa, np. emisji pyłów (PM10 i PM2,5), dwutlenku

10 <http://ryszardtadeusiewicz.natemat.pl/191987,internet-rzeczy> [dostęp: 1.10.2017].

11 Pew Research Center, (2017), <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet> [dostęp: 1.10.2017].

siarki, dwutlenku azotu etc., co związane jest z zanieczyszczeniem powietrza, mocą i zużyciem energii elektrycznej, energochłonnością procesów produkcyjnych etc., ale także śledzenia kontaktów przedsiębiorstwa z konsumentami. Monitoring jest jednak tylko punktem wyjścia do dokonania procesów produkcyjnych i tylko jednym z elementów cyfrowego DNA przedsiębiorstwa. Kolejnym istotnym elementem jest analityka danych z monitoringu procesów w całym łańcuchu logistycznym i definiowanie nowych funkcji produktów na podstawie tych danych i wreszcie definiowanie – na podstawie analizy danych o doświadczeniach konsumentów związanych z interakcją z przedsiębiorstwem – nowych usług dodatkowych dla konsumentów (konceptcja *Customer Experience*), co odzwierciedla przekonanie, że istotniejsze w kontaktach z konsumentami jest tworzenie wartości dla przedsiębiorstwa, i wykorzystanie wskaźników takich jak *customer live value* (CLV) zamiast tworzenia wartości dla samych konsumentów.

Możemy więc stworzyć mapę drogową do BI złożoną z dziesięciu następujących kroków:

1. Przewartościowanie tradycyjnych ról przedsiębiorstwa i na tej podstawie budowanie pożądanego obrazu idealnego przedsiębiorstwa w oparciu o przeprowadzoną analizę porównawczą;
2. Zaangażowanie konsumentów oraz innych interesariuszy w proces transformacji cyfrowej przedsiębiorstwa;
3. Unikanie odizolowanych rozwiązań – spojrzenie na inteligentne przedsiębiorstwo z szerszej perspektywy, wykraczającej poza stosowanie najlepszych praktyk;
4. Motywowanie do podejmowania inicjatyw przez pracowników, korzystania z innowacyjnych modeli biznesowych i realizacji innych działań na rzecz budowy inteligentnego przedsiębiorstwa;
5. Wdrożenie kompleksowej strategii zarządzania danymi w przedsiębiorstwie i stworzenie cyfrowych platform danych;
6. Stworzenie laboratoriów innowacyjności w celu promowania inspirowanych ekosystemów;
7. Zapewnienie bezpieczeństwa danym w całym cyklu ich życia;
8. Włączenie operatorów infrastruktury ICT w projekty dotyczące projektowania, finansowania i implementacji rozwiązań BI w przedsiębiorstwie;
9. Uzyskanie wsparcia dla idei cyfrowej transformacji przedsiębiorstwa i stworzenie systemu monitoringu i oceny inicjatyw podejmowanych w tym zakresie;
10. Ustanowienie komitetu sterującego programem transformacji cyfrowej przedsiębiorstwa i stworzenie stosownego systemu planowania dla tego programu.

Pytania kontrolne

1. Wymień poziomy piramidy informacyjnej. W jaki sposób mogą być one wspomagane przez systemy informatyczne?
2. W jaki sposób analizowane są wskaźniki efektywności? Co to są granulacje informacyjne?
3. Czym się różni mapa strategii od zrównoważonej karty wyników?
4. Co to jest System BI? Wymień jego najważniejsze cechy.
5. Omów najważniejsze typowe elementy Systemu BI.

Studium przypadku

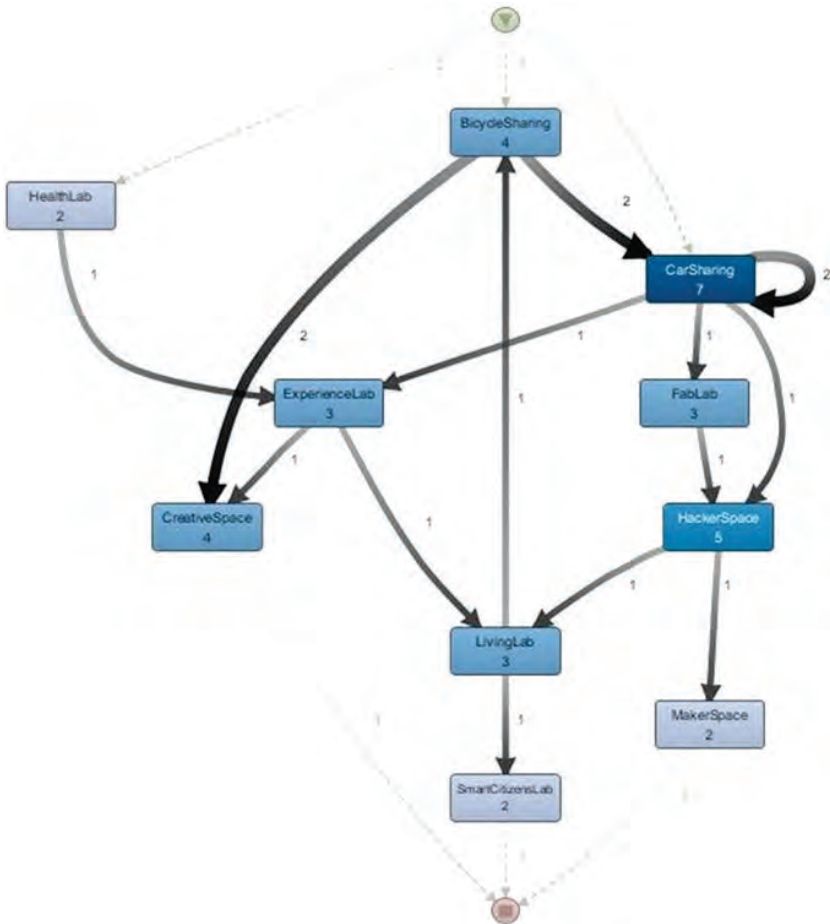
Przypadek 1

Przedstawione poniżej studium przypadku zostało opracowane dla problemu wyboru i hierarchizacji przedsięwzięć w hipotetycznym mieście na podstawie informacji dostarczanej przez system *Business Intelligence*.

Miasto zamierza przeprowadzić inwestycje związane ze strategią transformacji cyfrowej oraz budową miasta inteligentnego i zrównoważonego. Zdecydowano, że zostanie przeprowadzona analiza benchmarkingowa. Do porównań wykorzystano dane zebrane metodami wywiadu gospodarczego z miast uwzględnionych w indeksie miast zrównoważonych (*Arcadis Sustainable Cities Index*¹²). Korzystając z metod eksploracji procesów, utworzono mapę procesu budowy modelowego miasta, wykorzystując następujące dane: identyfikator instancji procesu (procesem jest droga do inteligentnego i zrównoważonego miasta rozumiana jako sekwencja typowych projektów realizowanych na rzecz idei inteligentnego i zrównoważonego miasta, instancją procesu są projekty jednego z wyróżnionych miast, identyfikator instancji procesu może być nazwą miasta), działania realizowane w trakcie realizacji procesu (projekty realizowane na rzecz idei inteligentnego i zrównoważonego miasta) oraz znacznik czasu (rok rozpoczęcia projektu). Dane te poddano analizie z wykorzystaniem algorytmu do mapowania procesów bazującym na algorytmie α [van der Aalst, 2011], otrzymując mapę procesu przedstawioną na rysunku 3.6.

12 www.arcadis.com/en/global/our-perspectives/sustainable-cities-index-2016 [dostęp: 7.11.2016].

Rys. 3.6. Mapa drogowa do inteligentnego i zrównoważonego miasta



Źródło: [Gontar, 2016].

Nietrudno zauważyć, że droga ta może rozpocząć się od wdrożenia innowacyjnego modelu biznesu „dostarczania funkcjonalności zamiast własności” [Gontar, 2016]. Inteligentne i zrównoważone miasto zapewnia wówczas usługi, które np. zaspokajają potrzeby komunikacji indywidualnej jego mieszkańców czy turystów, nie wymagając od nich fizycznego zakupu i utrzymania środka komunikacji. Dotyczy to zarówno rowerów, skuterów, jak i samochodów, w tym samochodów elektrycznych.

Koszty posiadania produktów fizycznych są ponoszone przez miasto, co może umożliwić mieszkańcom/turystom dostęp do wcześniej nieosiągalnych ze względu na koszty produktów. Druga widoczna innowacja

w zrównoważonym modelu biznesu, rozpoczynająca drogę do inteligentnego i zrównoważonego miasta, to „przyjmowanie funkcji wóldarza (Stewarda)”, tzn. podejmowanie przez miasto współpracy z mieszkańcami i innymi interesariuszami w celu zapewnienia im długoterminowego zdrowia i dobrobytu. Temu celowi ma służyć tworzenie żywych laboratoriów w obszarze zdrowia (*HealthLab*), które mają testować rozwiązania dostarczające korzyści środowiskowe i społeczne. Dopiero w dalszej kolejności wprowadzane są rozwiązania wspierające przedsiębiorczość.

Mając zdefiniowany wzorzec lidera na drodze do inteligentnego i zrównoważonego miasta, porównano otrzymaną mapę drogową z inicjatywami zrealizowanymi do tej pory przez miasto i stwierdzono, że zrealizowano już szereg przedsięwzięć z zakresu „dostarczania funkcjonalności zamiast własności” oraz „przyjmowania funkcji wóldarza”. Postanowiono więc, że miasto skoncentruje się na projektach wspierających przedsiębiorczość. Nowe rozwiązania, wspierające modernizację przemysłu i tworzenie nowych struktur przemysłowych w inteligentnym i zrównoważonym mieście (zidentyfikowane w analizie benchmarkingowej), bazują na pomysłach zrodzonych w *Massachusetts Institute of Technology* (MIT). Są to centra fabrykanckie (*FabLab* oraz podobne do nich *CreativeSpace* i *HackerSpace*), których celem jest wprowadzenie nowych technologii produkcyjnych do społeczności miejskiej. Każdy *FabLab* powinien być wyposażony w urządzenia, które – zgodnie z pierwotną ideą – umożliwiają wyprodukowanie praktycznie wszystkiego. *FabLaby* zobowiązane są do udostępniania swoich projektów oraz programów edukacyjnych, tworząc znaną w środowisku informatycznym otwartą społeczność. Najnowszym pomysłem, realizowanym w drodze do inteligentnego i zrównoważonego miasta, są żywe laboratoria (*LivingLaboratory*) testujące na żywym organizmie miejskim rozwiązania zgodne z modelem „maksymalnej efektywności materiałowej i energetycznej”, „gospodarki zasobami w układzie zamkniętym”, „zastępowania rozwiązań miejskich odnawialnymi alternatywami”.

Analiza benchmarkingowa doprowadziła do wytypowania szeregu projektów do realizacji (tab. 3.6).

Tabela 3.6. Portfel potencjalnych projektów do realizacji w inteligentnym i zrównoważonym mieście

| Projekt |
|--|
| Żywe laboratorium architektury energooszczędnej i energii odnawialnej (demonstracyjny/szkoleniowy inteligentny budynek, pełniący funkcję laboratorium dla Smart Grid Analytics), a jednocześnie rewitalizacja ceglanej fabryki |
| Żywe laboratorium miasteczka uniwersyteckiego |
| Miasto jako żywe laboratorium środowiska smart, zbudowane wokół Smart Factory |

| Projekt |
|--|
| Żywe laboratorium inteligentnej produkcji miejskiego samochodu elektrycznego |
| Budowa miejskiego środowiska do osobistej produkcji, wyposażonego w drukarki 3D oraz inne urządzenia i oprogramowanie niezbędne w mikrofabryce |
| Żywe laboratorium centrum miasta typu smart |
| Projekt idealnego miasta wykorzystującego rozwiązania smart |

Źródło: na podstawie [Gontar, 2016].

Zdecydowano, że w ocenie zdolności do osiągnięcia przez miasto równowagi pomiędzy celami ekonomicznymi, środowiskowymi i społecznymi wykorzystane zostaną wskaźniki KPI kojarzone z SDG 11¹³ (tab. 3.7).

Wartościowanie projektów inwestycyjnych wspomagających strategię zrównoważonego rozwoju i analiza ich wartości ze względu na efekty ekonomiczne, społeczne i środowiskowe przyjmuje formę analizy wielokryterialnej, np. analizy obwiedni danych (DEA) czy analizy hierarchicznej problemu (AHP). Inne propozycje dodają do analizy systemy z logiką rozmytą oraz algorytmy ewolucyjne. Możliwe jest zastosowanie także innych metod wielokryterialnych.

Wybór projektów jest w naszym przypadku prosty. Przełożenie celów strategicznych na konkretne działania miasta wymaga zdefiniowania składu portfela projektów. Wiele spośród zaproponowanych projektów odnosi się do idei żywego laboratorium. Koncepcja żywego laboratorium wyrosła z paradygmatu „inteligencji otoczenia” (*Ambient Intelligence*, AmI) opracowanego pierwotnie w korporacji Philips i przyjętego przez Grupę ISTAG (*Information Society Technologies Advisory Group*) pracującą dla Komisji Europejskiej. Stosownie do propozycji ISTAG w inteligentnym otoczeniu funkcjonować będą inteligentne interfejsy wbudowane w urządzenia codziennego użytku (w domach, biurach, samochodach i miastach), co spowoduje powstanie sieci inteligentnych urządzeń, które będą wspólnie pozyskiwać, przetwarzać i przysyłać informacje, dostosowując w sposób automatyczny swoje działanie do wymagań poszczególnych osób. Żywe laboratoria miały testować te urządzenia w warunkach operacyjnych. Z punktu widzenia środowiska uniwersyteckiego ciekawym przykładem żywego laboratorium jest projekt *MyCampus* zrealizowany na uniwersytecie *Carnegie Mellon*, gdzie przetestowano rozwiązania inteligentnego otoczenia w codziennym życiu kampusu uniwersyteckiego.

13 SDG11. Zrównoważone miasta i społeczności (uczynić miasta i osiedla ludzkie bezpiecznymi, stabilnymi, zrównoważonymi oraz sprzyjającymi włączeniu społecznemu).

Tabela 3.7. Złożoność decyzji a szczebel zarządzania

| Cel | KPI | Obliczenia |
|---|--|---|
| Zmodernizowanie śródmieścia, aby stało się ono miejscem do zamieszkania przez pracowników nowo powstałych przedsiębiorstw | Odsetek pracowników nowo powstałych przedsiębiorstw zamieszkałych w śródmieściu | NPS (nowa populacja w śródmieściu): = $100 \times$ # pracowników nowo powstałych przedsiębiorstw, którzy zamieszkali w śródmieściu / # pracowników nowo powstałych przedsiębiorstw |
| Modernizacja systemu transportu miejskiego oparta na produkowanym w mieście samochodzie elektrycznym | Odsetek mieszkańców, którzy mają dogodny dostęp do stacji ładowania samochodów elektrycznych (w sensie odległości) | Mieszkańcy z dostępem do stacji ładowania (w%): = $100 \times$ populacja z wygodnym dostępem / populacja miasta |
| Zrównoważona urbanizacja w śródmieściu | Wskaźnik zużycia gruntów do stopy wzrostu populacji w śródmieściu | Wskaźnik wzrostu populacji WWP = $\text{LN} (P_n / P) / y$ P – całkowita liczba ludności w śródmieściu w pierwszym roku realizacji programu P _n – całkowita populacja w śródmieściu w ostatnim roku realizacji projektu y – liczba lat między okresami pomiarowymi |
| Ochrona dziedzictwa | Wydatki ogółem (publiczne i prywatne) na mieszkańca wydane na rewitalizację fabryk | Udział wydatków na ochronę dziedzictwa w budżecie: $B_{di} = B_d, i / B_i$ B _d – procent rocznego budżetu przewidzianego na utrzymanie dziedzictwa kulturowego i naturalnego w roku i B _d – całkowita kwota rocznego budżetu przewidzianego na utrzymanie dziedzictwa kulturowego i naturalnego w roku i B _i – całkowita kwota rocznego budżetu publicznego w roku i |
| Zmniejszenie niekorzystnego wpływu miasta na środowisko | Odsetek stałych odpadów miejskich regularnie gromadzonych odprowadzanych | Udział regularnie zbieranych i odprowadzanych odpadów = Re / O Re – regularnie gromadzone i odprowadzane odpady O – całkowita ilość odpadów miejskich |

| Cel | KPI | Obliczenia |
|---|--|--|
| Powszechny dostęp do przestrzeni zielonych w mieście | Odsetek obszaru zabudowanego miasta, który jest zieloną przestrzenią do użytku publicznego | Udział zielonej przestrzeni w powierzchni zabudowanej miasta |
| Zrównoważone budownictwo | Wsparcie finansowe na budownictwo zrównoważone | Całkowita pomoc netto miasta na budownictwo zrównoważone |
| Swobodny dostęp do zasobów miasta | Odsetek mieszkańców z dostępem do podstawowych zasobów | Odsetek mieszkańców, którzy mają dostęp do podstawowych zasobów (wody, kanalizacji, usług gospodarowania odpadami stałymi, nowoczesnej i odnawialnej energii oraz transportu publicznego i dróg) |
| Poprawa jakości życia w mieście poprzez zwiększenie recyklingu w przedsiębiorstwach z branż niebezpiecznych zlokalizowanych w mieście | Odsetek odpadów poddanych recyklingowi z branż niebezpiecznych | Ankiety i spisy kierowane do przedsiębiorstw z branż niebezpiecznych |

Źródło: opracowanie własne na podstawie [Stair, 1992].

Należy zwrócić uwagę na to, że przyjęta metoda oceny i wartościowania projektów zgodna z zasadą rozwoju zrównoważonego zakłada równorzędność celów ekonomicznych, społecznych i środowiskowych. Na portfel projektów składają się projekty pogrupowane ze sobą w taki sposób, by uzyskać portfele, które realizują jak największą liczbę równorzędnych celów. Z jednej strony zatem chcielibyśmy otrzymać taki skład portfela projektów, który odnosi się do jak największej liczby celów, ale z drugiej strony chcielibyśmy mieć w portfelu takie projekty, które są najbardziej efektywne we wspomaganiu celów zrównoważonej turystyki.

Analiza efektywności jest przedmiotem badań wielu naukowców. Zaczynem były prace Pareta, Debreu, Koompansa, Leibensteina i Farrella, który zapoczątkował prace nad analizą efektywności organizacji. Do pomiaru efektywności projektów stosowane są różne metody, które pozwalają uzyskać dane stanowiące podstawę do podejmowania decyzji o wyborze i uszeregowaniu projektów (ich hierarchizowaniu). Zazwyczaj są to metody oparte na finansowych miernikach oceny projektów (wartości bieżącej netto, NPV i wewnętrznej stopie zwrotu, IRR), metody oparte na miernikach wartości projektów dla akcjonariuszy (na przykład metoda wartości dodanej dla akcjonariuszy: *Shareholder Value Added*, SVA) oraz metody równoważenia portfela, które pozwalają na

osiągnięcie celów strategicznych organizacji poprzez realizowany portfel projektów (macierz zgodności projektów ze strategią).

Analiza efektywności projektów innowacyjnych przeprowadzona w niniejszej książce wykorzystuje metodę programowania liniowego Data Envelopment Analysis (DEA), określaną w języku polskim zazwyczaj jako analizę danych granicznych lub analizę obwiedni danych, opracowaną w 1978 roku przez W. Coopera, E. Rhodesa i A. Charnesa na podstawie funkcji produktywności M.J. Farella rozumianej jako relację wartości otrzymanych rezultatów projektów pozostających w zależności od ich wkładu w realizację celów strategii zrównoważonej turystyki do wartości nakładów inwestycyjnych.

Tabela 3.8. Portfel projektów do realizacji w inteligentnym i zrównoważonym mieście po przeprowadzonej analizie

| Projekt |
|--|
| Projekt idealnego miasta wykorzystującego rozwiązania smart |
| Żywe laboratorium miasteczka uniwersyteckiego |
| Budowa miejskiego środowiska do osobistej produkcji, wyposażonego w drukarki 3D oraz inne urządzenia i oprogramowanie niezbędne w mikrofabryce |
| Żywe laboratorium architektury energooszczędnej i energii odnawialnej (demonstracyjny/szkoleniowy inteligentny budynek, pełniący funkcję laboratorium dla Smart Grid Analytics), a jednocześnie rewitalizacja ceglanej fabryki |
| Żywe laboratorium centrum miasta typu smart |
| Miasto jako żywe laboratorium środowiska smart, zbudowane wokół Smart Factory |
| Żywe laboratorium inteligentnej produkcji miejskiego samochodu elektrycznego |

Źródło: na podstawie [Gontar, 2016].

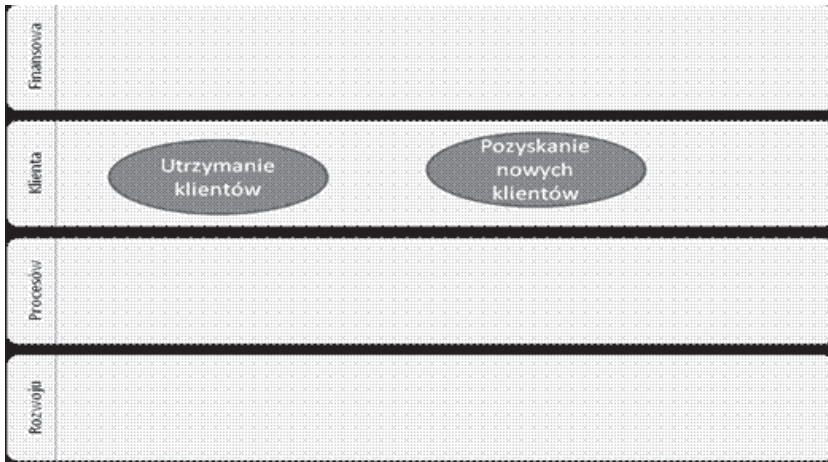
Przypadek 2

Przedstawione poniżej studium przypadku zostało opracowane dla problemu zarządzania sprzedażą kawy i słodkich przekąsek w hipotetycznych punktach sprzedaży znajdujących się na terenie uniwersytetu.

Uniwersyteckie centrum innowacji uruchomiło punkt sprzedaży kawy i słodkich przekąsek jako żywe laboratorium przedsiębiorczości akademickiej (punkt prowadzony przez studentów) i zamierza przeprowadzić inwestycje związane z utworzeniem sieci takich punktów sprzedaży w innych budynkach uniwersytetu. Zdecydowano, że w ramach żywego laboratorium przedsiębiorczości akademickiej zostanie opracowany i przetestowany system *Business Intelligence* monitorujący

efektywność funkcjonowania punktów sprzedaży. Ustalono, że istotne są dwa cele strategiczne: utrzymanie klientów i pozyskanie nowych klientów (rys. 3.7)¹⁴ oraz jeden wskaźnik efektywności (KPI): wartość klienta (wartość relacji z klientem) w perspektywie długookresowej (tab. 3.9).

Rys. 3.7. Mapa strategii w ujęciu klasycznym z perspektywami: finansową, klienta, procesów i rozwoju dla punktów sprzedaży kawy i słodkich przekąsek



Źródło: opracowanie własne.

Przyjęto najprostszą formułę wartości klienta zaproponowaną przez Gupta i Lehmana [Gupta, 2005], która może być rozumiana jako wartość bieżąca wszystkich przyszłych zysków generowanych przez klienta [por. Doligalski, 2010].

$$CLV = \sum_{i=1}^n \frac{m_i}{(1+i)^i},$$

gdzie

CLV to wartość klienta (ang. *Customer Life Value*),

m_i to zysk na kliencie w okresie i (dla uproszczenia obliczeń w przykładzie przyjęto, że wykorzystany zostanie przychód ze sprzedaży),

n to przewidywana długość cyklu życia klienta.

¹⁴ Widzimy zatem, że mamy prostą sytuację, w której rozpatrujemy tylko perspektywę klientów oraz tylko dwa niezależne od siebie cele strategiczne w tej perspektywie: utrzymanie klientów oraz pozyskiwanie nowych klientów.

Tabela 3.9. Metryki efektywności w perspektywie „klientów”

| | Cele | KPI |
|----------------------|--------------------------------|-------------------|
| PERSPEKTywa KLIENTÓW | (1) Utrzymanie klientów | – wartość klienta |
| | (2) Pozyskanie nowych klientów | |

Źródło: opracowanie własne.

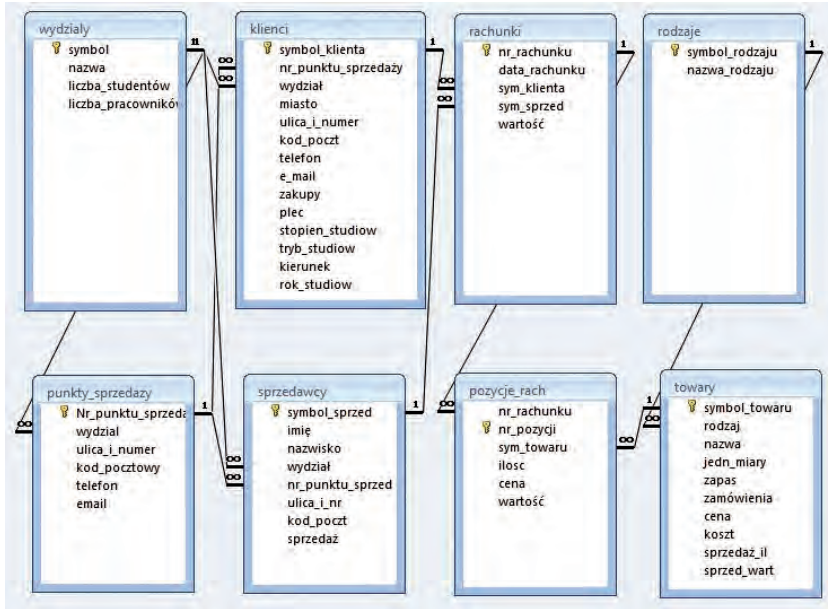
Zdecydowano, że przeprowadzona zostanie analiza benchmarkingowa. Do porównań wykorzystano dane zebrane z jednorocznej eksploatacji uruchomionego już punktu sprzedaży. Klientom (ograniczo- no rozważania do studentów, pomijając pracowników uniwersytetu) w uruchomionym już punkcie zaproponowano kartę stałego klienta, która umożliwiała identyfikację podstawowych danych charakteryzujących klienta: płeć, stopień studiów (np. licencjackie, magisterskie), tryb studiów (np. stacjonarne, niestacjonarne), kierunek i rok studiów. Mając zatem numer karty stałego klienta, można było powiązać informacje o sprzedaży z informacjami na temat klientów.

Wykorzystamy schemat bazy danych z rozdziału „Bazy danych”. Na potrzeby naszego przykładu dokonamy w nim małych zmian. Nasza baza danych powinna obejmować następujące encje:

WYDZIAŁY (zamiast KRAJE z bazy oryginalnej)
 PUNKTY_SPRZEDAŻY (zamiast BIURA z bazy oryginalnej)
 SPRZEDAWCY
 KLIENCI
 TOWARY
 RODZAJE towarów
 RACHUNKI (zamiast FAKTURY z bazy oryginalnej)
 POZYCJE rachunków

Relacyjny schemat bazy „Sprzedaż kawy i słodkich przekąsek” wygląda następująco (rys. 3.8):

Rys. 3.8. Schemat bazy „Sprzedaż kawy i słodkich przekąsek”

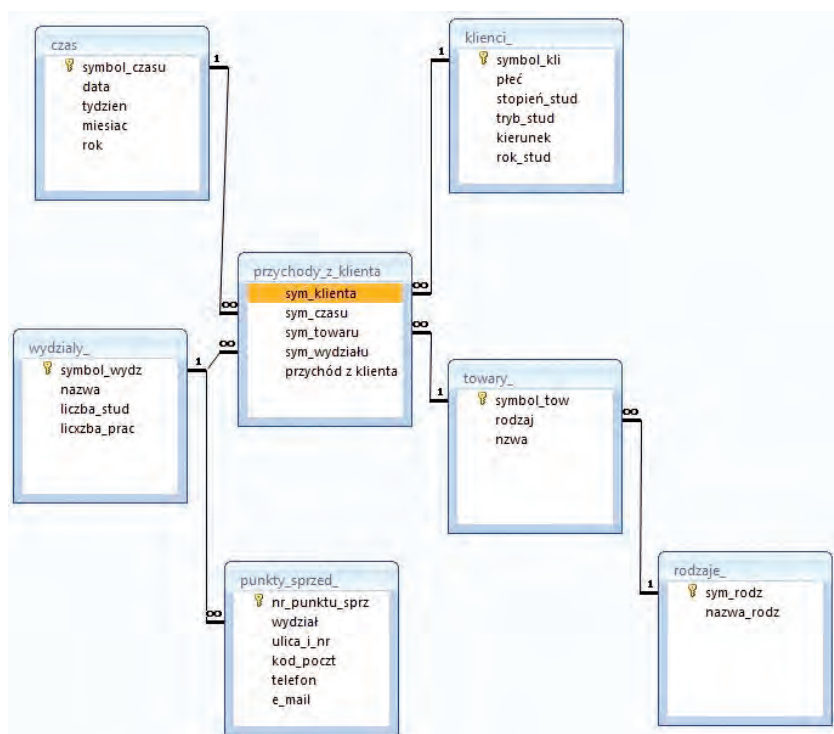


Źródło: opracowanie własne.

Wykorzystując metodę symulacji zdarzeń dyskretnych, należy użyć zestaw sztucznie tworzonych zdarzeń dla jednorocznej eksploatacji uruchomionego punktu sprzedaży. Dane o sprzedaży oraz dane o klientach należy zapisać w bazie SQL-owej „Sprzedaż kawy i słodkich przekąsek”. Dysponując tymi danymi, chcielibyśmy obliczyć KPI odnoszące się do efektywności sprzedaży (wspomniana wyżej wartość klienta) w następujących wymiarach: czasowym, towaru oraz klienta. Informacje dotyczące KPI są wprowadzane okresowo (raz dziennie) do bazy SQL-owej rozumianej jako hurtownia danych oraz do kostki OLAP-owej. Pamiętając sposób wyliczenia KPI, wiemy, że hurtownia danych powinna zawierać informacje o dziennym przychodzie ze sprzedaży poszczególnych klientów. O tym, w jaki sposób informacje mają być zapisane w hurtowni danych, decyduje definicja ziaren (granuli). Ponieważ chcemy monitorować przychód ze sprzedaży w sensie czasowym, towaru oraz klienta, ustalamy np., że najniższy poziom szczegółowości to przychód dzienny z pojedynczego klienta, z konkretnego punktu sprzedaży, z każdego towaru. Musimy zatem obliczyć na koniec dnia te wartości na podstawie danych ze sprzedaży (danych z bazy danych „Sprzedaż kawy i słodkich przekąsek”) i wprowadzić je do hurtowni da-

nych. Na wyższych poziomach szczegółowości przychód ze sprzedaży w tygodniu dla grupy towarów, dla studentów określonego kierunku studiów etc. obliczymy, tworząc agregaty w samej hurtowni na etapie ładowania danych do hurtowni lub analizy danych. Należy więc wykonać najpierw działania typu ETL (*extract, transform, load*) – pobrać dane z bazy danych, zagregować je dla każdego klienta i dla każdego towaru oraz zapisać je w hurtowni danych. Możemy wykorzystać skrypty SQL-owe, o których mowa w rozdziale „Bazy danych”, ponieważ wykorzystamy bazę SQL-ową. Schemat płatka śniegu hurtowni danych „Wartość klienta w perspektywie długookresowej wygląda następująco” (rys. 3.9):

Rys. 3.9. Schemat płatka śniegu hurtowni danych „Wartość klienta w perspektywie długookresowej”



Źródło: opracowanie własne.

W hurtowni danych (i w kostce OLAP) przechowujemy zatem informacje o dziennych przychodach ze sprzedaży z każdego klienta. Pamiętając o tym, że informacja, która raz trafiła do hurtowni danych, nie jest z niej usuwana, można w każdej chwili, stosując np. język SQL (do hurtowni danych) czy MDX (do kostki OLAP), uzyskać informacje o przychodach

ze sprzedaży w kolejnych dniach (tygodniach-miesiącach) etc., aktualną wartość każdego klienta dla danych historycznych (mamy możliwość regulacji szerokości ziarna/granuli, możemy więc np. obliczyć wartość klienta oddzielnie w odniesieniu do kawy i do słodkich przekąsek).

Język MDX (MultiDimensional eXtensions) został opublikowany w roku 1999 w firmie Microsoft jako część specyfikacji OLE DB for OLAP (ODBO) i został zaimplementowany po raz pierwszy w komercyjnym produkcie w systemie Microsoft OLAP Service 7.0. Obecnie stosowany jest także w innych systemach BI (np. Oracle Hyperion Essbase, SAS OLAP Server, SAP Netweaver BI i in.). Wewnętrznie program Microsoft Excel korzysta z języka MDX przy obsłudze tabel przestawnych. Język MDX został zaprojektowany w celu tworzenia pól wyliczalnych w kostce OLAP oraz tworzenia zapytań w środowisku OLAP (w wielowymiarowym modelu danych). Przykładowe proste zapytania języka MDX przedstawiamy poniżej.

```
SELECT  
[Measures].[przychod_z_klienta] ON COLUMNS  
FROM [Wartość klienta w perspektywie długookresowej];
```

```
SELECT  
[Measures].[przychod_z_klienta] ON COLUMNS15,  
towary.nazwa.Members ON ROWS  
FROM [Wartość klienta w perspektywie długookresowej];
```

Analiza OLAP-owa (z wykorzystaniem języka MDX) nie da nam odpowiedzi na pytanie, jaka jest wartość klienta (studenta) w perspektywie całego okresu studiów. Tu musimy sięgnąć do metod analizy predykcyjnej. Jeśli chcemy obliczyć przewidywaną wartość klienta na koniec okresu studiowania¹⁶, należy zastosować jedną z metod eksploracji danych (analizę szeregów czasowych). Możemy też dodać kolejny wymiar do analizy i obliczyć wartość klienta, uwzględniając nie tylko wymiar czasu, ale też wymiar towaru. Otrzymalibyśmy w ten sposób np. wartość klienta w odniesieniu do kawy 'latte' i odrębnie w odniesieniu do kawy 'americano'.

Poniżej zamieszczono przykładowe instrukcje DMX tworzące prognozy przychodów z klientów kupujących kawę 'latte' lub 'americano' w perspektywie najbliższych 3 tygodni.

```
CREATE MINING STRUCTURE [Wartość klienta]  
(
```

15 Wiedząc (z charakterystyki klienta), że klient będzie na studiach określoną liczbę lat (wartość n ze wzoru na CLV), możemy obliczyć całkowitą wartość każdego klienta.

16 Jw.


```

[Data] DATE KEY TIME,
[nazwa] TEXT KEY,
[Przychod_z_klienta] LONG CONTINUOUS
)
GO
ALTER MINING STRUCTURE [Wartość klienta]
ADD MINING MODEL [Wartość klienta dla kawy]
(
  [Data],
  [nazwa],
  [Przychod_z_klienta] PREDICT
)
USING Microsoft_Time_Series (FORECAST_METHOD = 'ARIMA')
WITH DRILLTHROUGH
GO
INSERT INTO MINING STRUCTURE [Wartość klienta]
(
  [Data],
  [nazwa],
  [Przychod_z_klienta]
)
OPENQUERY(, Wartość Klienta w perspektywie długookresowej ,,
,SELECT
  [Data],
  [nazwa],
  [Przychod_z_klienta]
FROM [vPrzychod_z_klienta]17 ORDER BY [Data]
,
)
GO
SELECT FLATTENED
PredictTimeSeries([Wartość klienta dla kawy].[Przychod_z_klienta],3)
FROM [Wartość klienta dla kawy]
WHERE [nazwa] = 'latte' OR [nazwa] = 'americano'

```

Mając prognozy przychodów z klientów kupujących kawę 'latte' lub 'americano' w perspektywie najbliższych 3 tygodni, możemy obliczyć wartość klientów z uwzględnieniem utworzonych prognoz. Wartość klienta można wykorzystać w ocenie projektów utworzenia sieci punktów sprzedaży w budynkach uniwersytetu.

¹⁷ Dla uproszczenia przykładu zakładamy, że w hurtowni danych utworzona została perspektywa wirtualna, która zawiera stosowne informacje.

Literatura

- Ambler S. (2011), *Agile Best Practices for Data Warehousing (DW)/Business Intelligence (BI) Projects*, www.agiledata.org/essays/dataWarehousingBestPractices.html [dostęp 7.12.2017].
- Beck K. i in. (2001), *Manifest Zwinnego Tworzenia Oprogramowania*, agilemanifesto.org/iso/pl [dostęp: 7.12.2017].
- Chaudhuri S., Staier J., Verma S., Lawton J., Miller J. (2010), *The Metrics Reference Model: A jumpstart for Business Intelligence initiatives*, „Cost Management”, September/October, s. 21–28 (Grupa robocza CAM-I Business Intelligence stworzyła Metrics Reference Model, aby pomóc organizacjom zbudować całościowe ramy pomiaru wydajności).
- Chmielewski M. (2009), *Wywiad biznesowy w korporacjach transnarodowych. Teoria i praktyka*, Wydawnictwo Adam Marszałek, Toruń.
- Doligalski T. (2010), *Pomiar wartości klienta – wartość życiowa klienta, LTV, CLV*, <http://www.doligalski.net/pomiar-wartosci-klienta> [dostęp: 16.08.2018].
- Elkington J. (2018), *25 Years Ago I Coined the Phrase „Triple Bottom Line”. Here’s why it’s time to rethink it*, „Harvard Business Review”, 25 June.
- Gontar B., Gontar Z. (2016), *Zrównoważona turystyka. Strategiczny wybór projektów ICT*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Gupta S., Lehmann D. (2005), *Managing Customers as Investments: The Strategic Value of Customers in the Long Run*, Wharton School Publishing, Pennsylvania.
- Kaplan R.S., Norton D.P. (1996), *The Balanced Scorecard: Translating Strategy Into Action*, Harvard Business Press, Brighton.
- Kaplan R.S., Norton D.P. (2001), *The Strategy Focused Organization*, Harvard Business School Publishing, Brighton.
- Kisielnicki J. (2008), *MIS systemy informatyczne zarządzania*, Wydawnictwo PLACET, Warszawa.
- Lin T.Y. (1997), *From Rough Sets and Neighborhood Systems to Information Granulation and Computing in Words*, European Congress on Intelligent Techniques and Soft Computing, September 8–12, 1997, http://xanadu.cs.sjsu.edu/~tylin/publications/paperList/101_34rnsigcw3.pdf [dostęp: 7.12.2017].
- Molski M., Łacheta M. (2007), *Przewodnik audytora systemów informatycznych*, Helion, Gliwice.
- Osterwalder A., Pigneur Y. (2012), *Tworzenie modeli biznesowych. Podręcznik wizjonera*, Helion, Gliwice.
- Stair R.M. (1992), *Principles of Information Systems – A managerial approach*, Thomson Publishing, Stamford.
- van der Aalst W.M.P. (2011), *Process Mining: Discovery, Conformance, and Enhancement of Business Processes*, Springer-Verlag, Berlin Heidelberg.
- Wolny W. (2004), *Metody odkrywania wiedzy w systemach Business Intelligence*, Prace Naukowe/Akademia Ekonomiczna w Katowicach, Systemy wspomagania organizacji SWO’2004.
- Zadeh L.A. (1979), *Fuzzy Sets and Information Granulation*, North Holland Publishing, Amsterdam.
- Zadeh L.A. (1997), *Towards a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic*, „Fuzzy Sets and Systems”, vol. 90.