

**Sebastian Andracki**

Uniwersytet Zielonogórski

## **DATA MINING JAKO NARZĘDZIE CONTROLLINGU DLA MAŁYCH I ŚREDNICH PRZEDSIĘBIORSTW**

### **1. Wstęp**

W wyniku prowadzonej na coraz szerszą skalę działalności gospodarczej przedsiębiorstw gromadzone są coraz większe ilości danych. Funkcjonalne związki pomiędzy nimi są często nierozpoznane lub w najlepszym przypadku – bardzo niepewne. Adaptując do opisu rzeczywistości gospodarczej wypracowane w teorii modele ekonomiczne, praktycy życia gospodarczego dopiero w ostatnim dziesięcioleciu stanęli przed szansą ich praktycznej weryfikacji. Postępujące procesy informatyzacji spowodowały rozwój elektronicznych metod gromadzenia i przetwarzania dużych zbiorów danych oraz doskonalenie metod ich eksploracji. Do takich metod zaliczane są: metody drążenia danych (ang. *data mining*), wizualizacji (OLAP), które z jednej strony dostarczyły bogatego instrumentarium, służącego odkrywaniu ukrytej w zbiorach danych wiedzy. Z drugiej zaś zrodziły problemy akceptacji wyników analiz uzyskiwanych za pomocą skomplikowanych metod teoretycznych.

### **2. Szanse i ryzyko wynikające ze stosowania *data mining***

Upowszechnianie się w coraz większym stopniu zintegrowanych rozwiązań informatycznych, np. klasy ERP (ang. *enterprise resource planning*) w sektorze małych i średnich przedsiębiorstwach (MŚP) stwarza pytanie o możliwość wykorzystania potencjału zgromadzonych w nich danych. Z jednej strony zauważane są korzyści wynikające z pracy ze zintegrowanym systemem, z drugiej zaś w szybszym tempie od przyrostu gromadzonych w archiwach systemowych danych rosną potrzeby użytkowników końcowych – zarówno najwyższego, jak i średniego szczebla – co do możliwości wykorzystania danych do generowania informacji zarządczej.

Pierwsze doświadczenia z wdrożenia standardowego oprogramowania klasy ERP proALPHA na polskim rynku pokazują, iż w większości przypadków pierwszy z trzech poziomów jakości informacji według trzypoziomowej koncepcji zachowania jakości danych opracowanej w Zakładzie Controllingu i Informatyki Ekonomicznej został osiągnięty<sup>1</sup>. Standardowe raporty i statystyki (bilans, rachunek zysków i strat, statystyki klientów, artykułów itp.), stanowiące dziś integralną część niemal każdego z oferowanych systemów, pozwalają zaspokoić większość podstawowych zarówno wewnętrznych, jak i zewnętrznych potrzeb informacyjnych. „Klienci” informacji po wdrożeniu systemu klasy ERP oraz zaspokojeniu potrzeb informacyjnych pierwszego poziomu coraz silniej wyrażają potrzebę dysponowania narzędziami do przygotowania i przetwarzania danych oferujących więcej niż tylko zdefiniowany „na sztywno” zestaw raportów i statystyk. Oczekiwania klientów informacji zmierzają w kierunku narzędzi, które pozwalałyby osiągać drugi oraz trzeci poziom jakości, a więc poziom „jakości zadowalającej” oraz „jakości zachwycającej” [4, s. 43]. Osiągnięcie tych kolejnych poziomów jest punktem wyjścia do opracowania i wykorzystania coraz bardziej zaawansowanych metod eksploracji danych, np. *data mining*.

W dalszej części przedstawiono poszczególne aspekty koncepcji zapewnienia jakości danych w sektorze MŚP na przykładzie modułu proALPHA Analyzer<sup>2</sup>.

## 2.1. Aspekty zapewnienia jakości danych

Zapewnienie jakości danych na poziomie zadowalającym podkreśla konieczność spełnienia wymagań dotyczących elastyczności w zakresie przygotowania informacji zarządczej i jej niezawodności [4, s. 44]. O elastyczności można mówić wówczas, gdy wiele zależnych od danej sytuacji przedsiębiorstwa potrzeb informacyjnych może zostać stosunkowo szybko i w różnorodny sposób (tabele, wykresy) zaspokojone. Bazują one w większości na wykorzystaniu tradycyjnych metod analiz, np.:

- kształtowania się obrotów i marży wybranych grup klientów,
- moralności płatniczej klientów,
- świadczeń gwarancyjnych.

W dalszej części artykułu przedstawiono przykład analizy moralności płatniczej klientów. Dane pozyskano z bazy systemu proALPHA. Następnie wczytano je do modułu proALPHA Analyzer. Przykład oparty został na rzeczywistych danych przedsiębiorstwa, które od kilku lat wykorzystuje standardowe oprogramowanie klasy ERP.

---

<sup>1</sup> Koncepcja zakłada występowanie trzech poziomów jakości danych: poziomu jakości wymaganej, zadowalającej i zachwycającej. Szczegółowo opisano ją w pracy [4, s. 41-74].

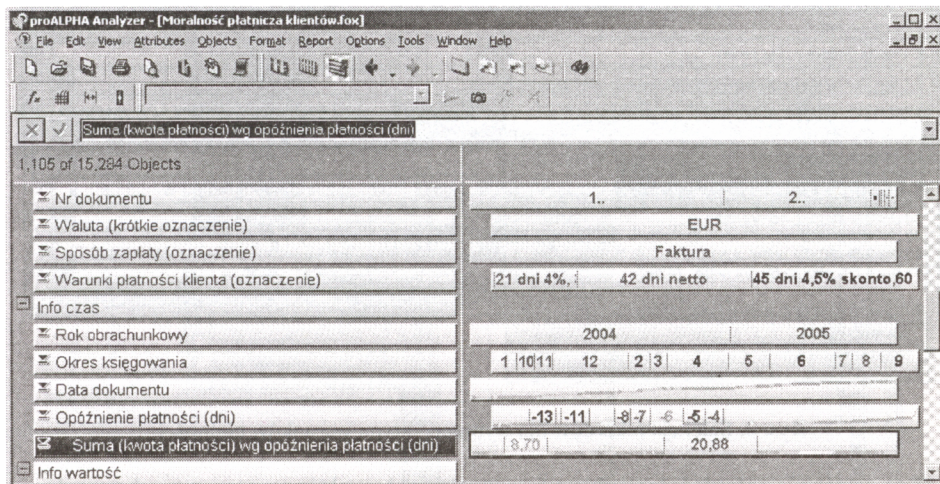
<sup>2</sup> proALPHA Analyzer to narzędzie bazujące na technologii OLAP, dostarczane wraz ze standardowym oprogramowaniem proALPHA. Stanowi rozszerzenie jego podstawowej warstwy sprawozdawczo-analitycznej.

### Przykład.

Poszukiwane są prawdopodobieństwa płatności klientów uwzględniające warunki płatności.

### Rozwiązanie.

Ponieważ w wariancie standardowym analizy dostępna jest tylko jedna (łączna) kwota płatności, należy utworzyć nowy atrybut *Suma (kwota płatności) według dni opóźnienia w płatności* w celu dokonania sumowania kwoty płatności według dni zwłoki (rys. 1):



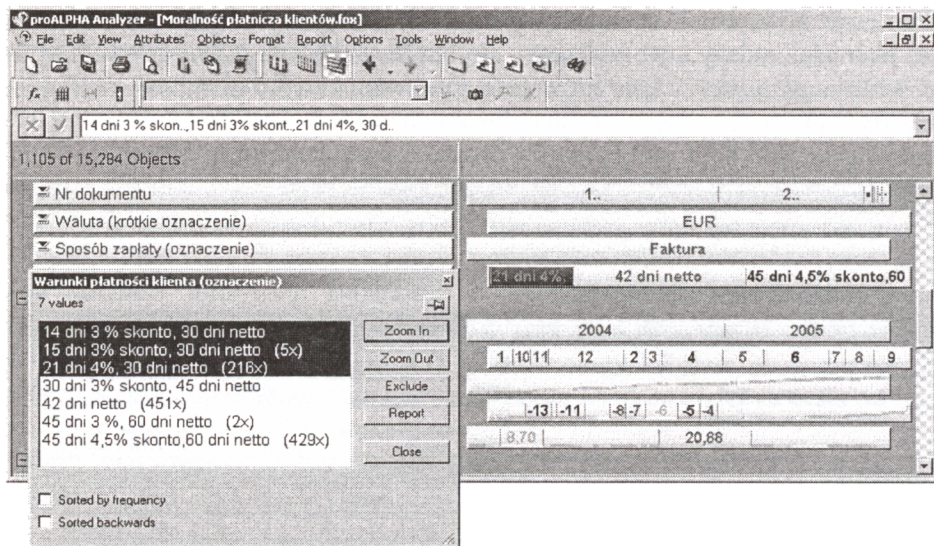
Rys. 1. Podsumowanie kwoty płatności według dni zwłoki

- operacja „zoom in” w polu *Warunki płatności klienta (oznaczenie)* dla tych warunków płatności, które mają tę samą liczbę dni netto (rys. 2),
- operacja „zoom in” w polu *Warunki płatności klienta* dla tych warunków płatności, które mają tę samą liczbę dni netto (rys. 2),
- sortowanie danych według atrybutu *Opóźnienie płatności (dni)*,
- przeprowadzenie kontroli jakości danych dla pożądaných atrybutów, np. dla atrybutu *Opóźnienie płatności (dni)*, *Suma (kwota płatności) według opóźnienia płatności (dni)*, ewentualne wykluczenie wartości niepożądanych (rys. 3).

Problematyka jakości danych w sensie ich poprawności oraz kompletności podjęta zostanie szerzej w dalszej części artykułu. Należy jednak mocno zaakcentować, iż możliwość wykrywania błędnych lub brakujących danych powinna być jednym z istotnych czynników wpływających na podjęcie decyzji o wyborze odpowiedniego narzędzia analiz.

Aspekt innowacyjny w ramach trzeciego poziomu zapewniania jakości danych podkreśla wykorzystanie zaawansowanych technik komputerowych (narzędzi wizualizacji, predefiniowanych raportów wywoływanych za pośrednictwem „jednego

kliknięcia”). Techniki te i narzędzia powinny bazować na zrozumiałych i akceptowanych przez adresatów informacji metodach analiz, których łączenie i wzajemne kombinacje umożliwiają uzyskanie nowych rezultatów analiz.



Rys. 2. Selekcja rekordów dla wybranego typu warunków płatności

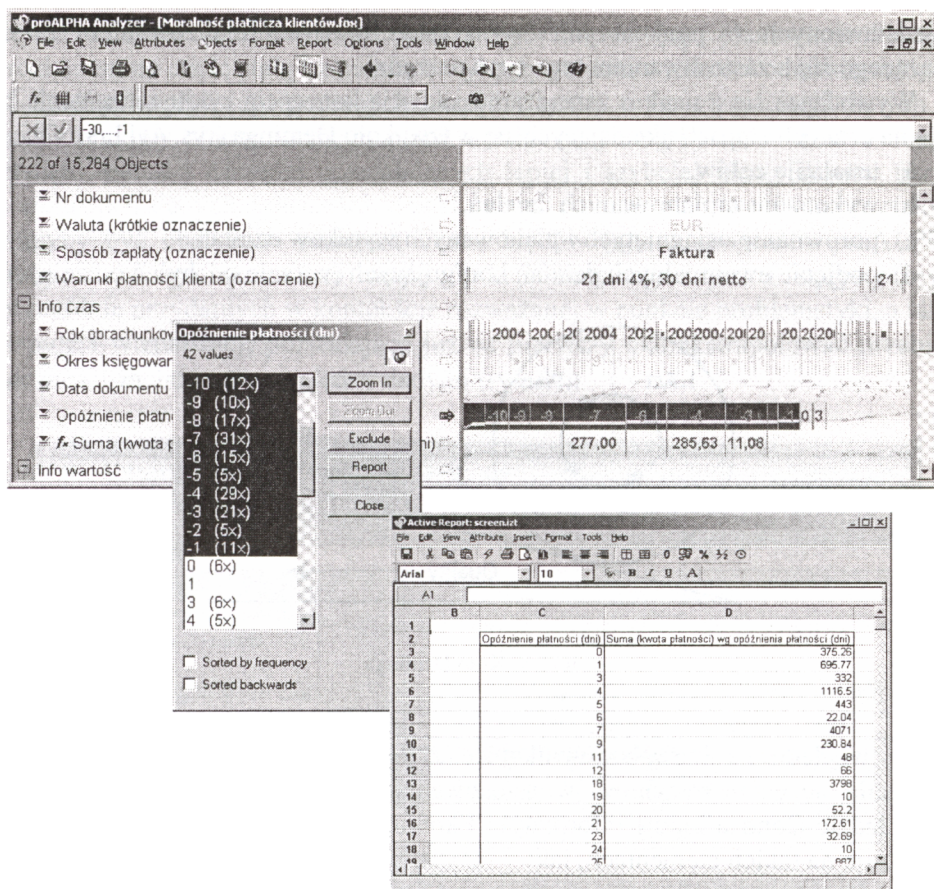
Zapewnienie jakości danych na poziomie zadowalającym oraz zachwycającym ograniczone jest możliwościami zapewnienia jakości danych na poziomie wymaganym. Spowodowane jest to tym, że dane gromadzone w archiwach przedsiębiorstw obciążone są błędami popełnianymi w trakcie realizacji podstawowych procesów dyspozycyjnych w przedsiębiorstwie. Źródłem tych błędów są:

- błędnie zdefiniowane lub niekompletne dane podstawowe (jedną z najczęstszych przyczyn pojawienia się problemów z jakością danych w przedsiębiorstwie),
- błędy popełniane w trakcie wprowadzania danych transakcyjnych,
- błędy będące następstwem operacji korygujących już wprowadzone dane (np. operacja storna) [5, s. 130].

Osiągnięcie wymaganego poziomu jakości danych ma duże znaczenie także dla innych obszarów zastosowań nowoczesnych metod komputeryzacji, np. APS<sup>3</sup>. Wykorzystanie tych metod prowadzić może do uzyskania znaczących efektów racjonalizatorskich w ramach przetwarzania już istniejących i generowania na tej podstawie całkowicie nowych informacji. Warunkiem koniecznym do osiągnięcia tego jest uzyskanie efektów racjonalizatorskich w ramach pierwszego, podstawowego poziomu – poziomu jakości wymaganej.

<sup>3</sup> *Advanced planning and scheduling.*





Rys. 3. Wyłączenie z analizy wartości niepożądanych

Zastosowanie technik drążenia danych *data mining* w sektorze MŚP wiąże się zwykle z istotną barierą. Może nią być nastawienie głównego kierownictwa (lub szefa) do gromadzenia, udostępniania, a także pozyskiwania nowych, bazujących na skomplikowanych technikach obliczeniowych informacji (problem akceptacji). Wewnętrzny system controllingowy w tych przedsiębiorstwach realizowany jest często tylko w uproszczonej formie. Archiwa danych charakteryzują się dużym stopniem rozproszenia, utrudniając tym samym ich eksplorację.

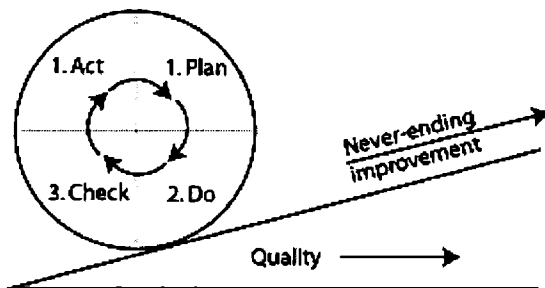
### 3. Zastosowanie narzędzi i metod *data mining* w kontekście cyklu zarządzania

Poszukiwanie możliwości zastosowania narzędzi i metod *data mining* w powiązaniu z cyklem zarządzania przedsiębiorstwem wiąże się z wykorzystaniem

przedstawionych w punkcie pierwszym szans oraz minimalizowaniem ryzyka wiążącego się z zastosowaniem tego typu narzędzi.

Wywodząca się z nauk o zarządzaniu jakością koncepcja cyklu zarządzania bazuje na czterech etapach układających się w tzw. koło Deminga (rys. 4) [1, s. 37]:

- a) ustalanie celów,
- b) wykonanie zaplanowanych działań,
- c) porównanie osiągniętych i planowanych wyników działań,
- d) podjęcie ewentualnych działań korygujących (w razie odchyłeń od założeń).



Rys. 4. Koło Deminga

Źródło: [2].

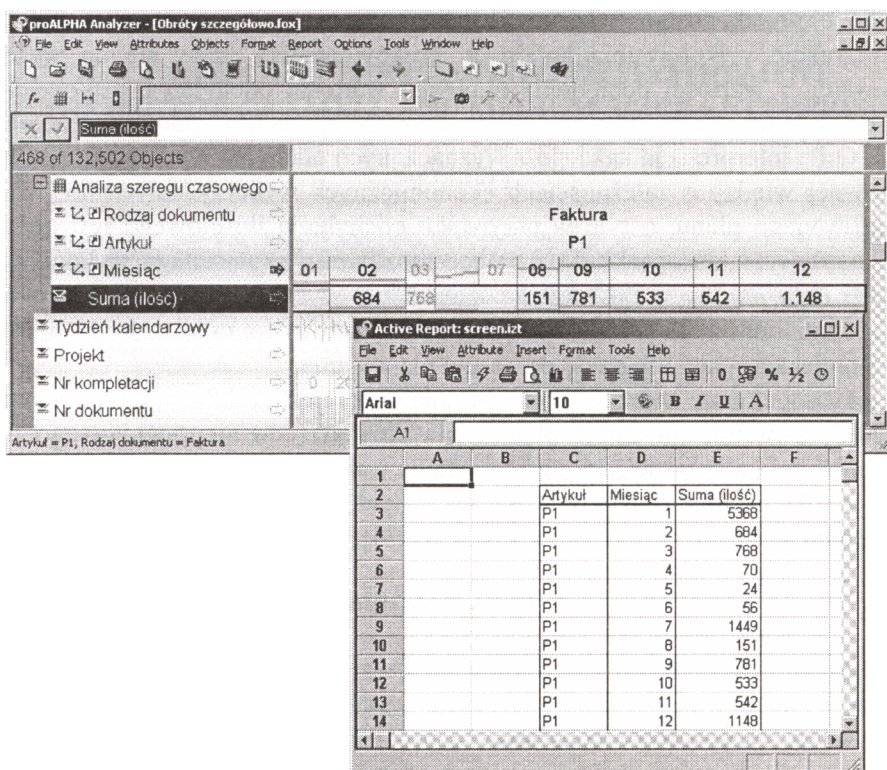
Ze względu na korzyści wynikające z jej stosowania koncepcja Deminga doczekała się wielu rozwinięć w zastosowaniach controllingowych. Opracowana w Zakładzie Controllingu i Informatyki Ekonomicznej koncepcja cyklu zarządzania bazuje na takich elementach, jak:

- a) zintegrowane planowanie działalności przedsiębiorstwa na bazie arkusza Microsoft Excel,
- b) rejestracja danych rzeczywistych w systemie ERP,
- c) analiza odchyłeń plan–działania wykonane w arkuszu kalkulacyjnym,
- d) korekta planu w sensie planowania kroczonego.

Ad a) Bazujące na modelach matematycznych zintegrowane planowanie działalności przedsiębiorstwa obejmuje swym zasięgiem przynajmniej najważniejsze obszary działalności, takie jak: sprzedaż, produkcja, zaopatrzenie/gospodarka materiałowa oraz rachunkowość. Istotą zintegrowanego planu, w odróżnieniu od tradycyjnych metod planowania, jest fakt, iż bazuje on na zastosowaniu zaplanowanych wielkości odniesień (np. wielkość sprzedaży), wykorzystanie których czyni proces planowania bardziej elastycznym, pozwalającym na uzyskanie większej przejrzystości powiązań pomiędzy elementami planu (np. powiązania między wielkością sprzedaży a wielkością przychodów, kosztów lub stanów magazynowych itp.) [6, s. 331]. Bardzo często do modelowania zależności pomiędzy wielkościami odniesienia a wynikami w zintegrowanym planie wykorzystuje się zależności liniowe. Może to prowadzić do powstawania błędów planowania, po-

nieważ zależności te wcale nie muszą być liniowe (np. planowanie kosztów z rozbiciem na część stałą i zmienną z wykorzystaniem wielkości odniesienia).

Dodatni wpływ nowoczesnych narzędzi analiz oraz wizualizacji danych na proces budowy zintegrowanego planu działalności mogłyby się wiązać z dostarczeniem metod i narzędzi pozwalających lepiej i szybciej identyfikować ukryte w archiwach danych zależności, dzięki czemu procesy planowania w przyszłości mogłyby być obciążone mniejszą liczbą błędów (np. w programie proALPHA Analityzer zaimplementowano wiele funkcji matematycznych i statystycznych szczególnie przydatnych w kontekście rozważanej w punkcie problematyki, a także narzędzia wizualizacji pozwalające już na pierwszy oka ekstrapolować przebiegi zależności zachodzące pomiędzy danymi (rys. 6)).



Rys. 5. Analiza szeregu czasowego. Wpływ sezonowości na poziom sprzedaży

Ad b) Rejestracja danych rzeczywistych następuje przede wszystkim podczas realizacji podstawowych procesów dyspozycyjnych i administracyjnych w przedsiębiorstwie. To właśnie dzięki realizacji tych procesów powstają coraz większe archiwa danych, których wykorzystanie jawi się jako ogromna szansa do zastosowania nowoczesnych metod i narzędzi analizy danych (por. punkt 1). W tym

kontekście podstawowego znaczenia nabierają następujące zagadnienia związane z rejestracją danych:

- problematyka jakości danych dotycząca wszystkich obszarów przedsiębiorstwa, w których stosowane są metod ilościowe (dla controllingu istotna jest jakość danych ekonomicznych wraz z wpływającymi na nie wielkościami odniesienia),
- potrzeba ujmowania danych rzeczywistych w taki sam sposób jak danych planistycznych (w celu zapewnienia realizacji analizy plan–działania wykonanej),
- problematyka kształtowanych kreatywnie danych z obszaru rachunkowości.

Ad c) Bazująca na koncepcji elastycznego rachunku kosztów analiza odchyłeń plan–działania wykonane dostarcza cenne informacje controllingowe, pod warunkiem że jej wyniki zostaną właściwie zinterpretowane.

W trakcie wykonywania kolejnych kroków analizy<sup>4</sup> dąży się do wyselekcjonowania z wartości odchylenia ogólnego wartości odchyłeń cząstkowych, w tym: odchylenia parametrów modeli w obszarze sprzedaży, zakupów, rachunkowości. O ile interpretacja każdego z wyznaczonych odchyłeń wymaga od analityka gruntownej wiedzy o zależnościach ekonomicznych wykorzystanych do budowy zintegrowanego planu, o tyle w przypadku odchylenia ilościowego jest ona najtrudniejsza (stąd też oczekiwania wobec możliwości skutecznego wykorzystania narzędzi *data mining* w obszarze są szczególnie wysokie). Dotychczas udało się wskazać kilka przyczyn powstania odchylenia ilościowego, m.in.:

- zmniejszającą się integrację procesów planowania,
- błędy popełniane w procesie planowania oraz rejestracji danych rzeczywistych (błędy powstałe na etapie planowania prowadzą do nieprawidłowego funkcjonowania zależności, np. wykorzystanie wartości indeksowanych do oddzielnego planowania rodzajów kosztów stałych na podstawie wartości rzeczywistych z okresów poprzednich) [4, s. 67-74].

#### 4. Podsumowanie

W artykule podjęto próbę przedstawienia warunków ramowych wykorzystania narzędzi i metod *data mining* do eksploracji archiwów danych przedsiębiorstw z sektora MŚP. Zasygnalizowano szanse i zagrożenia wykorzystania tego typu narzędzi, ze szczególnym uwzględnieniem tych obszarów controllingu składających się na tzw. cykl zarządzania, dla których połączenie tradycyjnych metod analiz (np. analiza odchyłeń plan–działania wykonane) z nowoczesnymi metodami eksploracji i wizualizacji danych może skutkować rozwojem nowych, lepszych technik analiz. Niniejsza praca opisuje wyniki początkowych rozważań autora nad tym kierunkiem badań.

---

<sup>4</sup> Techniczny aspekt przygotowania analizy omówiony został szczegółowo w pracy [4].

---

## Literatura

- [1] Ciechan-Kujawa M., *Rachunek kosztów jakości*, Oficyna Ekonomiczna, Kraków 2005.
- [2] [http://pl.wikipedia.org/wiki/Cykl\\_Deminga](http://pl.wikipedia.org/wiki/Cykl_Deminga).
- [3] <http://www.zcie.uz.zgora.pl/materialy.html>.
- [4] Kluge P.D., *Zapewnienie jakości danych jako aktualne zadanie kadry kierowniczej*, Oficyna Wydaw., Siedlce 2005.
- [5] Kluge P.D., Kuźdowicz P., *Zarządzanie jakością danych przy zastosowaniu rozwiązania klasy ERP w małym i średnim przedsiębiorstwie produkcyjnym*, Warszawa 2004.
- [6] Kluge P.D., Kuźdowicz D., Kuźdowicz P., Orzeszko P., Ostrowska-Huba A., *Technologia komputerowo wspomaganego controllingu w MŚP w warunkach nowej bazylejskiej umowy kapitałowej – Basel II. Uwarunkowania sukcesu przedsiębiorstwa w gospodarce opartej na wiedzy*, SUCCESS 2004, materiały z konferencji naukowej, Kazimierz Dolny 2004, Lublin 2004 .

## DATA MINING AS THE CONTROLLING INSTRUMENT IN THE SECTOR OF SMALL AND MEDIUM-SIZED ENTERPRISES

### Summary

The article discusses the chances and risks of implementation of modern analytical methods and tools (OLAP and Data Mining) to retrieving the huge data archives. Application of new opportunities refers to a concept of management cycle. The problem of data quality in the context of three lever data quality approach is also presented.