Progress of organisational data mining in South Africa

M Hart

Department of Information Systems, University of Cape Town, Private Bag, Rondebosch 7700, SOUTH AFRICA

RÉSUMÉ

Ce papier décrit trois études, à caractère principalement qualitatif, échelonnées sur une période de cinq ans, ayant pour but d'analyser des données exploitées dans diverses organisations Sud Africaines. L'objectif était d'acquérir une plus grande compréhension des problèmes auxquels les participants eurent a faire face dans le processus d'exploitation de données. Des interviews très élaborées furent utilisées. L'accent fut plutôt mis sur les problèmes d'origine commercial et non pas sur les aspects technologiques et algorithmiques. Durant le cours de nos recherches, des progrès concrets furent réalisés et un modèle pour l'exploitation de données fut proposé.

MOTS-CLÉS: fouille de données, entrepôt de données, informatique décisionnelle.

ABSTRACT

This paper describes three largely qualitative studies, spread over a five year period, into the current practice of data mining in several large South African organisations. The objective was to gain an understanding through in-depth interviews of the major issues faced by participants in the data mining process. The focus is more on the organisational, resource and business issues than on technological or algorithmic aspects. Strong progress is revealed to have been made over this period, and a model for the data mining organisation is proposed.

KEYWORDS: data mining, data warehouse, business intelligence.

1 INTRODUCTION

The data mining literature is regularly enhanced by technological and algorithmic advances in data mining and knowledge discovery approaches, and many new examples are given of strategically and profitably useful industrial, government and social applications. There is however far less published research that discusses the general organisational use of data mining, its enablers, barriers, and success factors, particularly in Africa.

This paper covers three different studies of data mining in large South African organisations, over a five year period. In so doing, it tracks the progress of data mining locally, and highlights the important issues prevalent at the time. One of the major initial reasons for these studies was to investigate what was actually being carried out in the data mining sphere locally, and to what extent the 'commercially publicised and marketing' activities were exaggerations or hype.

Data mining suffers from a nomenclature problem. Much of the public and the corporate world think of it as a superset of decision-support, business intelligence or analytics, rather than as a subset. This appears to be largely due to data mining being a catchphrase

Email: M Hart mhart@commerce.uct.ac.za

used loosely by marketers and the popular press for a broad analytical area. Definitions also vary widely in the academic literature.

Nemati and Barko[1] define organizational data mining (ODM) as "leveraging data mining tools and technologies to enhance the decision- making process by transforming data into valuable and actionable knowledge to gain a strategic competitive advantage". They continue however, "ODM covers a wide array of technologies, including but not limited to e- business intelligence, data analysis, SQL, customer relationship management (CRM), eCRM, EIS, digital dashboards, information portals etc."

In this paper a much narrower view of organisational data mining is used, similar to that of[2]: "Data mining is the process of selection, exploration, and modelling of large quantities of data to discover regularities or relations that are at first unknown, with the aim of obtaining clear and useful results for the owner of the database" It is seen as one component of business intelligence, business analytics or decision support, and separate from OLAP, query and reporting, etc. It draws on machine learning, statistical analysis, modelling techniques and database technology.

After a short review, the paper outlines the general methodology used for the three studies, then covers each study sequentially, devoting more detail to the most recent one. Learning points are made, and

a model for the data mining organisation suggested. The paper finishes with conclusions and recommendations. It is hoped that the research covered may be useful to organisations elsewhere who are embarking on data mining activities.

2 BACKGROUND

In view of the format of the paper and of this volume, little space will be devoted here to a literature review of data mining in general, and its technical aspects will not be discussed. Readers should find additional sources in other papers in this specialised volume. The focus in this paper is on the actual progress made with data mining in South African organisations.

2.1 Applications

Data mining is applied in various industries from investment management and retail solutions to marketing, manufacturing and health care. Standard applications include customer profiling, market basket analysis, fraud detection, churn management and credit scoring[3, 4, 5, 6, 7, 8]. [9] identifies the following major users:

Retailers, Banks, Telemarketing, Airlines,

and Telephone companies the primary focus has been profiling and targeting of customers in (CRM).

Credit card companies to approve credit card applications, make purchase authorisation decisions, analyse cardholders' buying behaviours and detect fraud.

Financial Services to build trading and risk models for developing investment strategies, carry out credit scoring and identify fraud patterns.

Recent published applications include product recommendations[10], criminal network analysis[11], supermarket customers[12], healthcare bad debt[13], outpatient revisit intention[14], and competitor analysis[15].

2.2 Progress of data mining generally

Despite these examples, data mining has found it far more difficult than OLAP and query tools to enter the corporate mainstream, tending to be a specialist application for skilled professionals. This is illustrated by Gartner's hype cycle[16] for maturity of business intelligence, which in general is expected to have significant growth and development for the next five years.

Data mining as a part of business intelligence is shown as having passed the peak of inflated expectations, and having reached the bottom of the trough of disillusionment. Of the twenty components in the business intelligence hype cycle, seven are expected to plateau within two years, nine within five years, and four in between five and ten years. Only data mining is expected to take more than ten years to reach maturity, while approaches like OLAP, business activity monitoring, corporate performance management,

dashboards, and analytical CRM are predicted to do so in less than half that time.

2.3 Barriers to data mining

Data mining success may be hampered by both technical and non-technical barriers, the latter often more significant[5]. Data mining requires skills not typically found in IT departments and data warehouse groups. "You need the skills of a statistician, of a computer scientist and of a business analyst to frame the problem in a way that makes sense."[5, 17, 18] highlight a communication gap between IT and business. For this reason cross-functional teams should interact closely with the business units to carry out data mining.

The technology used in data mining applications ranges from highly automated tools, that give the user little to adjust, to manual systems that require considerable technical knowledge. Data quality issues such as inconclusive data, error contamination, data uncertainty, and database modification arise[19], and prior to the commencement of data mining tasks, data cleansing is needed[19, 18]. This normally forms a significant part of a data warehousing project.

2.4 Some points for successful delivery

The success of data mining projects depends as much on business knowledge as on technological knowledge, yet business users are often left uninformed. [17, 20] note that, as with most projects, commitment and funding from business units should be established from the outset. The initiative also needs to be orchestrated by the business units. Putting the right team together is more important than the purchase of particular tools[18].

2.5 Data mining research in South Africa

Published research on data mining in the South African context has, to date, been limited. Companies have acknowledged the need to obtain value from the data stored in their data warehouses[5, 21, 22]. A study performed on prevalence of data mining in South African companies[23] found that only 9% of the sample studied used data mining techniques. Other published South African research[24, 25] generally covers more technical aspects of data mining.

3 RESEARCH METHODOLOGY

The paper examines various aspects of the usage of data mining over a five year period. In three separate studies (1997, 1999 and 2002), informed users of data mining from large organisations were interviewed. Expert opinion on important issues was sought, rather than selecting random samples to estimate statistical prevalence. In each study, in-depth interviews were carried out in a number of organisations that were household names in South Africa, and in most cases in the top three by size in their industries. Participants generally came from the IT department, from

marketing or from financial areas, and there were often more than one per organisation. Consultants from well-known organisations with data mining expertise were sometimes included in a study and / or to assist piloting.

The interviewees should not be regarded as representative in a positivistic, statistical sense, but were selected because, given their size and industry types, they could be expected to be more knowledgeable than most about data mining. In addition the research approach was largely qualitative, with an interest in exploring the phenomenon of data mining. Little research into this field has been carried out in South Africa, and most has been technology-directed or quantitative.

In many cases, but not all, the same organisation was interviewed in a subsequent study. The sampling design falls under Menard's[26] description of non-quite-longitudinal designs, closer to a revolving panel design than a repeated cross-section design or repeated survey[27]. This is quite acceptable, as no attempt is made at serious quantitative measurement of cohort analysis, temporal changes and trends, intra-individual change, causal relationships etc. The data is more open ended and qualitative, and while a few questions may be constant in each study, the issues being examined varied as the use of data mining matured. Interviews were almost always tape recorded, with participants' agreement, and text analysis later used to extract key themes and verify notes taken.

In some ways the research is an examination of the ongoing adoption and diffusion of data mining in the South African corporate world. It deliberately does not use the quantitative approach of TAM or its extensions[28, 29], the theories of[30] and others, or the recent comprehensive Unified Theory of Acceptance and Use of Technology (UTAUT) model[31]. Instead, a series of largely semi-structured interviews are used to build up a rich picture of the local usage of data mining, with a focus that progresses from one study to the next. The aim is to look at the use of data mining approaches, and related barriers and enablers, in the corporate environment, rather than the algorithmic, statistical and technological aspects that are well covered in other research. Advantages and disadvantages of semi-structured interviews are discussed in[32].

4   THE FIRST STUDY - 1997

This was very much an exploratory study. The term data mining had been used in literature and by companies and vendors for a few years, but there appeared to be limited examples locally of its use. Was there more hype than substance? Although great strides were being made with parallel processors, data warehouses and software applications, how much was data mining really entering the corporate world? What were users' and managements' perceptions of it, and in what ways did it differ from the statistical, query and reporting techniques in use for many years?

Motivation	Number
Information for strategic decisions	13
Better support for all decision makers	13
Understand customers better	8
Cheaper & more accurate reporting	5
Improve data quality	3
To do Data Mining	3
Other	3

Table 1: Motivation for implementing a data warehouse

In the mid-1990's the same 'nappies and beer' case was being constantly used as the prime example of data mining, and it was not clear how much real data mining was taking place in local companies. This was reminiscent of the early days of competitive advantage systems, when American Airlines and American Hospitals were invariably used to extol the benefits of such systems, and few other cases were mentioned. Data mining also appeared to be a convenient new marketing term often referring merely to long- established statistical techniques, fairly standard queries or on-line analytical processing (OLAP).

Of 24 large companies contacted in the three main South African cities of Cape Town, Johannesburg and Durban, 15 agreed to interviews. Seven were retailers, five were from financial institutions, two were in telecommunications, and one was an international consulting firm. All were household names, and were selected either because their data warehousing or data mining activities had been publicised in the business or computer press, or because they appeared likely to be frontrunners in data mining usage. In interviews a number of set questions were asked, but interviewees were encouraged to share their views and comments more broadly on a range of aspects. It soon became apparent that it was important to clarify each interviewees' own definition of the term data mining early in the interview.

Perhaps the major finding, which soon became apparent, was that most organisations were still implementing, modifying, or reconstructing a data warehouse or data mart. Twelve claimed to have a production data warehouse, but in most cases they were still completing implementation. Their focus of attention at that stage was on storing the data, not on methods of mining it. A summary of their motivations for implementing a data warehouse is shown in Table 1, revealing that only three were due to data mining.

While ten organisations were extracting data with OLAP or report generators, only four were doing any data mining. Of these four, three were pilot financial and retail projects that had been going for less than a year, and only one, in telecommunications, had lasted between one and two years. All had fewer than five end users, in each organisation from the IT Department, and also in two cases, from Marketing. In three cases models had been built by the IT Department (once with Marketing help), and in the fourth by outside consultants.

The most frequently cited reason for using data

mining was to improve marketing. A few had achieved some success in that area through customer profiling or grouping of their products. The other main reasons stated for using data mining tools were identification of new opportunities, and cost reduction, while increasing revenue received only two mentions. Some current difficulties raised were that it was not easy to apply the knowledge discovered, due to time and budget constraints as well as lack of a clear direction. Current processing power for many was also inadequate, and cost justification for data mining projects was generally more difficult than for others.

However, the general response to data mining was positive. More than two thirds of organisations interviewed planned to initiate data mining projects or expand their use of them in the future, while the balance planned to continue at their present level.

5   THE SECOND STUDY - 1999

Two years later it was appropriate to embark on another study of data mining progress in South Africa[5]. A detailed literature survey uncovered many issues that were regarded as important internationally in the data mining arena. These were roughly grouped into four categories : business issues, technical / technology issues, resource issues and organisational issues. These were considered to be factors (enablers or inhibitors) that could influence the growth of data mining locally.

Twenty four large companies in the Western Cape Province were contacted, and fifteen agreed to participate: three retailers, three in telecommunications, four insurance companies, two in entertainment and three management consultancies with expertise in the data mining area. Interviews contained structured questions on the importance of each of the issues from the literature, together with a few additional issues added during pilot studies. Interviewees were asked to rate each issue as very important, important or not important, and were also able to add their own suggestions to this list. Much of the one to two hours used in interviews was also allocated to open discussion of broader data mining aspects. Caution should be used in extrapolating these views beyond the Western Cape, although all companies have nationwide branches

All but one of the fifteen companies felt data mining was strategically important or very important, and that use of the data should be a management decision, not an IT one. Five organisations were currently involved with data mining projects, and the other ten were planning to use data mining tools in the future. Although the proportion of interviewees using data mining regularly did not seem to have changed much in the two years since the previous study, knowledge and awareness of the issues had improved. A selection of interviewee quotes illustrates some views at the time of the second study.

"The data mining project will fundamentally change the way the company operates" (large retailer)

"The concept is not new but tools are" (accounting firm)

"One in ten companies realise the value of data mining" (DM consultancy)

"Worry about relying on results 'that the tool spits out" (Internet publisher)

"People who sell data mining do not really know what it does - they just 'hype it up" (insurance company)

"Wouldn't outsource a data mining project" (various)

"It takes too long to get results. Management are a bit scared of it - if you get technical they back off". (insurance company)

"Information is inaccurate as people lied on their application forms" (entertainment cmpany.)

"Many firms say they are using data mining when in fact they don't know what it is" (marketing manager)

"Impractical for managers to do it themselves - a time-consuming technical exercise" (financial organisation)

"A key point is it informs you of connections and patterns in your data you would never have thought to ask about" (retailer)

Altogether 47 issues were suggested by the researchers, and by the interviewees themselves, spread almost equally over the four categories. As a guide to the interviewees' overall assessment of importance of different issues, Table 2 lists the top 20, ranked by the number rating them as very important.

Table 2 illustrates that, even discarding the outdated Y2K issue, not one technical or technology issue (TI) entered the top ten, while a range of resource (RI), organisational (OI) and business (BI) issues dominated. There was a common perspective that a team approach, combining statistical, business and IT skills, was the best, that data mining was strategic and should be business driven. Potential problems included skilled resources and management support.

From these ratings and the discussion it was clear that there was still some confusion about what data mining was, and that other tools such as OLAP and on-line query were in far wider use. Yet data mining was regarded as an effective, potentially strategic, approach that could enhance competitiveness, and those not currently using it were all planning to do so in future.

6   THIRD STUDY - 2002 : OBJECTIVES AND APPROACH

After three years there was an expectation that data mining should have matured further, and that its progress should be further investigated[33].

6.1   Research objectives

The primary objective of the research was to gain further understanding into the implementation and

Rank	Group	Issues	Not Imp	Imp	Very Imp
1	RI	Teamwork is essential	0	0	15
2	BI	The need for data mining is business driven	0	2	13
3	RI	Skilled resources are a problem	1	1	13
4	OI	Strategic advantage for large companies to use DM	1	2	12
5	OI	Trend to be internationally competitive	1	2	12
6	BI	Need a sponsor and management support	3	2	10
7	BI	Competitive pressure to data mine	4	2	9
8	OI	Company infrastructure is a problem	5	1	9
9	BI	Y2K issue	6	0	9
10	RI	Resources needed to set up or maintain the DM project	1	6	8
11	OI	Impact of departmental integration on better DM results	3	4	8
12	TI	Data warehouse needed	5	2	8
13	BI	Initiative arises from the marketing dept	6	1	8
14	BI	Gap between IT and business	4	4	7
15	TI	Lack of understanding of the term DM	5	3	7
16	BI	Cost as an issue	3	6	6
17	TI	Data integrity is a problem	5	4	6
18	RI	Internal issues are a problem	6	3	6
19	OI	Size of organisation	7	2	6
20	TI	Years that data has been accumulated	8	1	6

Table 2: Most important issues in 1999 study

progress of data mining in South African organisations. It was decided to split this into six secondary objectives:

1. To determine the issues that led to the implementation of data mining in South African companies.
2. To determine the main data mining activities performed by South African organisations.
3. To determine the factors influencing the implementation of data mining in South African organisations.
4. To establish the current difficulties experienced with data mining implementation in South Africa.
5. To establish the organisational issues facing South African companies that are currently implementing data mining.
6. To evaluate the organisations' actual performance of data mining against their expected performance.

6.2   Research approach

In order to obtain a richer analysis the author again opted to conduct semi-structured interviews with a selection of participants currently engaged in data mining. Purposive sampling was used to select several large companies located in the main business centres of Johannesburg and Cape Town that experience high data throughput (and thus store large volumes of data), and are currently implementing data mining projects. The final sample consisted of eight companies currently performing and implementing data mining and three that provide data mining-related services. The eight were classified as follows: Banking (2), Insurance (2), Retail (2), Retail Financial Services (1), and one as "Confidential", due to a concern that

they would be recognisable if their industry was given. Those interviewed came from the following business areas: knowledge management, IT, marketing, project management centre, client and intermediary analysis, and outsourcing. This gave varying perspectives into the different uses of data mining across the organisation. Interviews averaging an hour and a half took place during July and August 2002, in Cape Town and Johannesburg, and for many interviews more than one respondent represented the company.

Companies interviewed are large organisations from South Africa's two biggest business centres, but it should be noted that the sample used may not fully reflect the current data mining situation in South Africa. The aim of the research was however to obtain a deeper insight into issues surrounding the use of data mining, not to estimate prevalence.

In interviews a prompt-sheet was used, split into the following areas: Organisation Information, Data Mining Project Initiation, Data Mining Project Development, Data Mining Practice within the Organisation, Future Plans for Data Mining, and Additional Comments. The participating companies were very receptive and generally willing to divulge more information than was initially requested. All interviewees authorised recording of interviews for later transcription, and appeared at ease with this process.

7   FINDINGS OF THIRD STUDY

Before discussing results related to the six objectives some comments on the understanding of data mining will be made. Some issues concerning the data itself will also be covered.

7.1 Current Understanding of Data Mining

All interviewees understood the underlying theory of data mining and defined the term as they perceived it, based on their experiences and backgrounds. One from the Service industry defined data mining as:

"Predicting patterns that are not readily seen through analytical queries."

A respondent in the Banking sector stated that: "Data mining is really looking for trends within your data that can answer all the questions that you didn't really think about, and seeing what value you can get out of that."

A similar definition was also put forward by a respondent from the Retail industry: "It's really looking for trends and patterns in the data in order to predict future behaviour."

Despite varying focal points a broad consensus emerged that data mining was customer-centric, born out of statistics, with desired outcomes generally of increased sales and profitability. Although the interviewees understood the potential of data mining, many outside their function did not. A Banking interviewee voiced his frustration with the queries requested by other departmental managers:

" . . . the type of work they do is nothing more than simple queries and extracts and the presentation of those results in terms of typical frequency analysis. So you have a person who is qualified in Stats get put into the organisation and they say, 'Can you tell me how many customers we have that are female? How many customers do we have in that age group? How many of them that are female are over this age?' And they start asking for a linear type solution to the problem."

7.2 The Role and Problems of Data

One organisation collected up to 80 million records a day and a bank in the sample amassed data for about 70 million cheque transactions a month from different locations. The most common forms of data collected in Retailing relate to customer profiling and management. Secondary information sources such as census data are also used by organisations within the Insurance and Retail Financial Services industries. Banks collect client data at various transaction points, and the value of collecting this is seen as more than simple data capture:

"The decision has been made that we're not archiving our data, but growing it."

Insurance firms collect important data from intermediaries, such as brokers, who form an integral element of their sales force. A project manager from this industry highlighted his organisation's approach to identifying what data is to be collected:

" . . . we actually went back to all the areas where information at the end of the value chain is used, and found out what they actually need, and then we reverse engineer . . . "

7.2.1 Data Quality

All companies indicated that their primary source of data for data mining purposes is their data warehouse, and external data is used minimally if at all. A participant consultant from the Service industry has filled a niche for services to resolve the problems of poor data quality and integrity. Their services include: Data Management, Business Intelligence, Data Analysis, Data Transformation, Data Cleansing and Data Conversions.

"We have found that clients' data is not as accurate as it is thought to be. Management is generating reports based on this, and you cannot make decisions based on data that is inaccurate. You therefore need to clean the data at root level to make intelligent decisions at the top level."

This underlines the substantial problems many organisations have with "dirty" data.

7.2.2 Ownership and completeness of data

Ownership was identified as problematic by participants within the Banking and Insurance industries, where those performing data mining are required to motivate their need for the data before gaining permission from other departments to mine it.

Insurance companies encounter difficulties in obtaining the latest client details, and manually capturing data received from brokers and agents. The advent of cash cards and loyalty cards has meant that retailers are now able to obtain valuable demographic and behavioural data of many of their cash buyers as well as credit customers.

7.3 Objective One: Data mining initiation

7.3.1 Business problem or opportunity ?

Table 3 depicts the problems and/or opportunities that triggered the sample organisations' data mining project initiatives, as identified in interviews. As can be seen, CRM was identified by seven companies as being a contributing factor to initiation. A respondent within the Retail industry explained their need for CRM:

"We've got all those people on the database, and without starting to look at how they are buying, when they are buying and what they are buying, there is a lost opportunity. It wasn't really a major problem, it was more of an opportunity."

While most of the triggers identified emerged internally, the respondents in the sample still cited organisational commitment as a problem that needs to be addressed. On balance, more opportunities than problems were cited.

7.3.2 State of development

Four of the eight companies implementing data mining were at the initiation or roll-out stages of their data mining projects. The remaining four were currently running multiple projects and thus found themselves at various phases of development across a range of

Problems	No. of orgs.	Opportunities	No. of orgs.
Lack of insight into their business	4	CRM	7
Rising costs	3	Customer retention	4
Isolated customer views	2	Exploiting data collected	4
Lack of business knowledge	2	Generate opportunities	3
Competition	1	Increase sales	2

Table 3: Problems and Opportunities Triggering Data Mining

data mining endeavours. The two Banking organisations boasted data mining projects dating back between five and ten years. The two Insurance companies have been operating their data mining projects for between one and five years. The four remaining companies were still grappling with scalability, data and testing issues in their initial projects.

7.4 Objective Two: Data mining activities

Seven interviewees indicated that their organisations adopt both the directed and undirected approach to data mining, depending on the business problem and desired outcome. A Banking interviewee noted that while his organisation uses both approaches, they are usually required to use the directed approach, which limits the potential of discovering hidden patterns and trends.

"We prefer using the undirected approach, but nine times out of ten, we're being asked to do something that would use the directed approach."

7.4.1 Uses of Data Mining

The major uses of data mining (in decreasing order of the numbers of times mentioned in interviews) are: CRM and direct marketing, Customer and product planning, Fraud detection, Credit scoring, Market basket analysis and Propensity modelling.

The two Insurance companies use data mining as a tool in designing their product offerings, and customer attraction and retention strategies. Fraud detection, credit scoring and market basket analysis are also used to further add value and reduce risk in these organisations. Propensity modelling is a technique adopted by an organisation in the Banking industry to identify the tendency of clients to save or spend. One interviewee noted that they use data mining to predict where to optimally position their ATM machines, so as to keep running costs to a minimum. They also pointed out that a rival bank is using data mining to predict the occurrence of bank robberies, and have realised cost savings.

A respondent from the Retail industry indicated a growing opportunity of using data mining to leverage relationships in the value chain.

"Take Company X as an example. They supply Product A and Product B. There are fantastic opportunities for them in this situation. We [as the retailer] can tell them who is buying Product A, but not buying any of Product B. With the prediction capability that we've got, we can go even further and say who of the other two million people in our database should buy Product B. So there is a win-win in this for the supplier and for us."

7.4.2 Data mining practice

The companies primarily use data mining for predictive purposes, largely aimed at forecasting customer buying behaviour and fraud detection. The use of association, classification and sequential patterns aims to add customer value by providing more personalised services and identifying their preferences. A Bank also uses credit risk modelling and principal component analysis to enhance other processes in its product design areas. Regression analysis and decision trees were the techniques most commonly used within the sample, and other techniques commonly adopted were cluster analysis, neural networks and visualisation. Four respondents indicated that their data mining activities have unveiled unexpected patterns in the data, and that they were able to use this knowledge in their business operations. This thus represents knowledge that would not ordinarily have been obtained without data mining and provides a basis on which the organisation may restructure its strategies.

7.4.3 Objective Three: Implementation determinants

7.4.4 Procurement decision

All eight of the companies initiated data mining projects in-house, and one of these has since outsourced the function. Retention in-house was attributed to the following reasons:

"sensitive information essential to the organisation's competitiveness"

"growth in the organisation's intellectual capital"

"As data mining is used to formulate organisational strategies, it is a continuous process within the organisation that requires long-term commitment"

7.4.5 Budget

A Retail Financial Services company, in the initial stages, is required to provide a return on investment to enable further expenditure. Two sample companies operated on a per-project approval basis, requiring approval of their motivation for each major resource spend before moving ahead. The remaining budgets range from R5 million to R40 million; the data mining spend is larger within the Banking and Insurance industries. No respondents had exceeded their data mining budgets, as the necessary procedures and controls were in place.

7.4.6 The need for data warehouses

Seven participants in the research implemented data warehouses in-house. The other is currently implementing its warehouse, outsourced to one of the Service companies interviewed, at an approximate cost of R55 million. Another Service company confirmed that development and implementation costs of a data warehouse range from R10-R100 million. Initial costs may be lower, but ongoing maintenance increases the cost considerably.

"That's why you'll find that a small company will not have a data warehouse - because they cannot afford it."

7.4.7 Skills perceived as important for data mining

The main requirements for effectively practicing data mining activities were given as a sound business understanding, followed by analytical skills and a statistical background. These were best combined in a team. All organisations currently practising data mining possess the skills to do this within the organisation. However, three respondents said they were understaffed, while one respondent noted the need to improve the current skill level within the organisation. One company interviewed practises employee skills-modelling to profile employee skills levels within the data mining function, and has defined three main roles within the data mining team: Statistician, Operations Analyst and Project Manager.

7.4.8 Objective Four: Current implementation and difficulties

7.4.9 Project development process

Those with IT backgrounds outlined the project development process as that of integrating information from the organisations' various source systems into a single view and transforming and redefining the data. Those from Statistical / Analytical backgrounds described it as commencing with the problem definition, getting a feel for the data and extracting the relevant variables, and then developing models around those variables. The IT function is responsible for ensuring a single, consistently defined view of the data. This is indicative of the necessity for alignment between the IT function and the data mining function in the organisation.

7.4.10 Difficulties experienced

Seven respondents using data mining techniques identified difficulties that they are experiencing with their methodologies and implementation processes:

- A lack of organisational acceptance.
- Explaining to others what is being done, due to a lack of global understanding of the data and the techniques.
- Insufficient data to effectively perform data mining.

- Data mining, as a solution, is not currently being exploited. "It's not being used to its full potential, and then senior management wants to know what the ROI has been."
- A lack of intuitive skills.
- Poor data quality and integrity in the source systems, which adversely affects the validity and credibility of the data mining outputs

7.5 Objective Five: Organisational issues

All eight companies cited their data mining projects as being initiated at a high management level: marketing management in three, other directors in another three, and the managers of the Knowledge Management and Credit Risk departments.

All companies claimed some degree of alignment between the IT department and the department driving the data mining process. One respondent regarded the alignment with the IT department as minimal (they were also experiencing problems with the aggregation of data from various strategic business units, which has amplified their data ownership problems). Four said they were aligned for operational purposes, and four for strategic planning purposes.

While all companies cited the customer as being the main beneficiary of data mining outputs, the department within the organisation that used the outputs to enable this value-adding process was largely marketing.

7.6 Objective Six: Performance evaluation

7.6.1 Initiator's perceived benefits

The benefits of data mining most commonly perceived by the initiating person or department prior to the implementation of data mining were (in descending order of mention): Understanding customers better, Increasing cross-selling, Enabling better decision-making, and Providing single view of customer. Other potential benefits included: Obtaining a better understanding of the data, Making projections, Becoming more cost-effective, Identifying leads, trends or needs to serve, Safeguarding against malicious practices, Improving the direct marketing response rate.

7.6.2 Current expected benefits of data mining

In contrast, the benefits that the organisations currently expect to result from data mining are: Improved marketing and sales, Customer profiling, Reduced costs, Increased competitiveness, Improved fraud detection, Effective risk management, Increased profit, Improved credit risk management and Customer retention.

7.6.3 Actual performance of data mining versus expectations

Five organisations were unable to rate the actual performance of their data mining projects as they stated that it is too early to gauge. The two Banks noted

that most data mining projects had exceeded expectations, but a few had under achieved. One company felt that their data mining project performances were fair but needed improvement, because of the number of assumptions required to compensate for the lack of quality data.

7.6.4 Direct improvements resulting from data mining implementation

Interviewees listed these direct improvements resulting from the use of data mining:

- Higher growth and increased turnover in certain business areas
- Reduced bad debts and healthier credit portfolios
- Improved sales leads
- Improved client differentiation, customer profiling and customer retention
- Reduced call-centre costs
- Better quality and more accurate information.
- Operational and Strategic advantages

7.6.5 Specific strategies resulting from data mining outputs

All organisations, except for one still in the initial stages of data mining implementation, have developed specific strategies (in Marketing, Operational, Risk Management, CRM, Business and Financial) using the outputs of their data mining projects. Three respondents feel their strategies have proved very successful in retaining customers, two that they have been fairly successful, while two others state it is too early to gauge. An Insurance interviewee articulated the organisation's perception of data mining with respect to their specific strategy formulation:

"It is one of the key elements of the organisation's vision. It is becoming successful. We have launched a number of strategies and business targets based on the information we are getting from data mining. The organisation's vision is around wealth creation, based on knowing the client and having a relationship with the client which is directly related to CRM which is founded on data mining."

7.6.6 Future plans

All eight companies implementing data mining expressed their intentions to continue and expand the use of data mining within the organisation in the future.

8 LEARNING POINTS FROM THE THIRD STUDY

Based on their experience with data mining implementation and activities, the respondents suggested several key learning points:

"Develop interfaces that will ensure that ground-level data is captured accurately and efficiently. This ensures that the data used for data mining activities is of a consistently high quality and a reliable source."
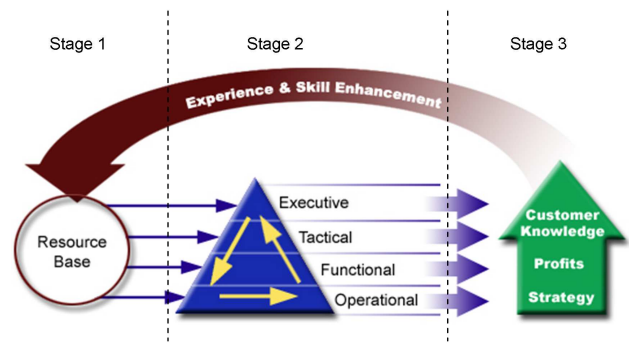

Figure 1: The Data Mining Organisation

"Management needs to recognize how much effort goes into the implementation of data mining projects, and their implicit buy-in needs to be obtained over time.: Cost is not an issue as much as a lack of commitment."

"Build simple models that are tested, to identify the issues that are going to be faced, before spending time developing complex models that are difficult to implement and time-consuming to refine".

"Rather than spending time developing models that will enable only a marginally competitive edge, it is more important to get the process right and resolve data integrity and skills issues".

"There is an inherent need to develop a culture within the organisation that understands the value of knowledge and the improved outputs that the cyclic process of data mining delivers".

9 THE DATA MINING ORGANISATION: A MODEL

Many aspects related to successful delivery of data mining in South African organisations have been covered in these three studies. It is suggested that key factors could be summarised in the form of the model in Figure 1.

By introducing the necessary processes and elements in this model the author is of the opinion that South African organisations would place themselves in a strong position to reap the benefits data mining offers. The model comprises three distinct "stages":

9.1 Stage 1

This area is the knowledge and resource base from which the various levels of the organisation are provided with information relevant to their specific functions.

9.2 Stage 2

This phase is indicative of the processes that should occur within the organisation. Awareness needs to be developed within the organisation with regards to the subtleties surrounding data mining. This is represented by the arrows within the organisation (represented by the triangle). The horizontal arrow illustrates where the responsibility of data quality, in-

tegrity and capture is assigned. The rising arrow illustrates the transformation of data to actionable information, then knowledge and ultimately decisions (taken at various management levels). The downward arrow represents the path on which the strategies, culture and objectives are passed down the organisation. This, in turn, determines the responsibility that needs to be assigned in order to ensure the quality and integrity of the data captured. This cyclical process thus ensures a dynamic, flexible orientation toward data mining within the organisation.

9.3 Stage 3

This stage indicates the organisation's realisation of their data mining activities in terms of results. Irrespective of the outcome (positive or negative), the organisation will be able to glean the necessary learning points, experience and ability from their efforts in order to contribute to their resource or knowledge base. This thus completes the process of a larger cycle that inculcates continuous learning upon the organisation in its approach toward data mining.

This model is focused on continuously developing a "Data Mining Culture" and assigning organisation-wide responsibility, supported by robust research and a practical resource base.

10 CONCLUSIONS AND RECOMMENDATIONS

The aim of this research was to provide an understanding of the progress of data mining in South African organisations, by outlining three qualitative, in-depth studies of large local companies who have been involved in that area.

10.1 Conclusions

The first exploratory study in 1997 showed that despite much comment in the press, often from vendors, relatively little real data mining was taking place. Companies were instead engaged in planning, installation and ongoing development and modification of their data warehouses. Most extraction of data from these warehouses or elsewhere was through basic query and reporting or OLAP tools, and few had used data mining as a reason to motivate for their data warehouse. Those engaged in data mining had only been doing so for a few months, mostly in pilot mode, and under five users. There was however overall agreement amongst those interviewed that they would embark more fully on data mining at a future stage.

The 1999 study showed far more awareness of a wider range of issues surrounding data mining, although the proportion of interviewed companies who were practising it strongly had not increased much. There was still much confusion amongst staff in the interviewees' companies as to the meaning of data mining. All interviewees agreed on its strategic value, and that it should not be managed purely by the IT function. Forty seven issues concerning data mining, including enablers and inhibitors, were rated according to their importance by those interviewed. It was notable that issues concerning technology or technical aspects of data mining did not make the top ten list. The issue that received unanimous backing was the need to undertake data mining as a team, combining sound analytical, business and IT skills. All organisations interviewed were planning data mining projects for the future.

By 2002 a major shift had taken place, with interviewees' organisations all having gained experience with data mining. They were able to describe a large and diverse number of practical applications, using different data mining approaches. Six research objectives were set for this study, and useful information was gathered from interviewees towards each of them. These covered data mining initiation, activities, implementation influences, implementation difficulties, organisational issues and performance evaluation. Current understanding of the term "data mining" and the role and problems of data were also explored. Added outcomes were a list of learning points, and a model designed to improve the practical application of data mining in the typical corporate organisation. This model of the "Data mining organisation" provides a framework that South African companies may use as a reference in establishing a culture that embraces data.

Companies currently performing data mining are successfully adopting a wide range of techniques, ranging from conventional to unique, tailored uses. The fact that some organisations are providing data mining-related services is indicative of further growth in the data mining industry in South Africa. Those embarking on it should be aware of the need to consider a set of important organisational, business and resource-related issues, as well as seeking appropriate technical solutions. While it is still a relatively niche area of business intelligence, its adoption, diffusion and success appear set to continue within South African organisations.

10.2 Recommendations

Practical recommendations are encompassed in the Data Mining Organisation model and the mechanisms it is equipped with. Key employees that hold positions across various levels of the organisation should be educated with the information pertinent to their specific functions. This would directly address some of the difficulties earlier identified. Another practical recommendation expressed by a number of the participating organisations is the establishment of a local benchmarking forum whereby organisations currently adopting data mining can further augment their data mining ability and knowledge by learning from others' experiences.

With respect to the proposed Data Mining Organisation model, research may be conducted to investigate the relevance and validity of the elements it identifies as success factors in data mining implementation. Furthermore, this study may be extended to

evaluate the extent to which these variables affect the success of data mining.

ACKNOWLEDGMENTS

REFERENCES

[1] H. R. Nemati and C. D. Barko. "Key Factors for Achieving Organisational Data Mining Success". *Industrial Management and Data Systems*, vol. 103, no. 4, pp. 282–292, 2003.

[2] P. Giudici. *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley & Sons, Ltd, 2003.

[3] K. H. Chye and C. K. L. Gerry. "Data Mining and Customer Relationship Marketing in the Banking Industry". *Singapore Management Review*, vol. 24, no. 2, pp. 1–27, 2002.

[4] M. S. Garver. "Using Data Mining for Customer Satisfaction Research". *Marketing Research*, vol. 14, no. 1, pp. 8–12, 2002.

[5] M. L. Hart, K. Davies, E.-J. Barker-Goldie and A. Theron. "Issues Affecting the Adoption of Data Mining in South Africa". *South African Computer Journal*, vol. 29, pp. 40–48, December 2002.

[6] E. Rafalski. "Using data mining / data repository methods to identify marketing opportunities in health care". *Journal of Consumer Marketing*, vol. 19, no. 7, pp. 607–613, 2002.

[7] J. Weir. "Data Mining: Exploring the Modern Corporate Asset". IT Knowledgebase, Auerbach Publications, CRC Press LLC, 1999.

[8] D. Yeo. "Is Data Mining Merely Hype?" IT Knowledgebase, Auerbach Publications, CRC Press LLC, 2000.

[9] L. Chen and T. Sakaguchi. "Data Mining Methods, Applications, and Tools". *Information Systems Management*, vol. 17, no. 1, pp. 65–70, 2000.

[10] D.-R. Liu and Y.-Y. Shih. "Integrating AHP and Data Mining for Product Recommendation Based on Customer Lifetime Value". *Information and Management*, vol. 42, no. 3, pp. 387–400, 2005.

[11] J. Xu and H. Chen. "Criminal Network Analysis and Visualisation". *Communications of the ACM*, vol. 48, no. 6, pp. 100–107, 2005.

[12] P. Lingras, M. Hogo, M. Snorek and C. West. "Temporal Analysis of Clusters of Supermarket Customers: Conventional vs Interval Set Approach". *Information Sciences*, vol. 172, no. 1–2, pp. 215–240, 2005.

[13] J. Zurada and S. Lonial. "Comparison of the Performance of Several Data Mining Methods for Bad Debt Recovery in the Healthcare Industry". *Journal of Applied Business Research*, vol. 21, no. 2, pp. 37–54, 2005.

[14] K. J. Lee. "A Practical Method of Predicting Client Revisit Intention in a Hospital Setting". *Healthcare Management Review*, vol. 30, no. 2, pp. 157–167, 2005.

[15] Y. P. Sheng, P. P. Mykytyn and C. R. Litecky. "Competitor Analysis and Its Defences in the E-Marketplace". *Communications of the ACM*, vol. 48, no. 8, pp. 107–112, 2005.

[16] F. Buytendijk. "The Business Intelligence Scenario: From Pressure to Performance". Gartner Symposium ITxpo 2005, 1-3 August, Cape Town, 2005.

[17] J. L. Butt., C. Rutstein, F. E. Gillett and S. Khawaja. "Turning Data into Dollars". Forrester Research, 2001. URL http://www.forrester.com.

[18] A. Linden. "Data Mining: Common Pitfalls and Solutions". Gartner Group, 1998. URL http://www.gartner.com.

[19] C. C. Chou and A. Y. Chou. "A Managers Guide to Data Mining". *Information Systems Management*, vol. 16, no. 4, pp. 33–41, 1999.

[20] A. Linden. "Data Mining Applications of the Next Decade". Gartner Group, 1999. URL http://www.gartner.com/DisplayDocument?id-299 858&acsFlg=accessBought.

[21] M. L. Hart, M. G. Jacobs, G. D. Jooste and J. N. Saulez. "Are South African Companies Mining their Data Warehouses?" Tech. rep., University of Cape Town, Department of Information Systems, 1997.

[22] H. L. Viktor and W. H. Motha. "Creating Informative Data Warehouses : Exploring Data and Information Quality through Data Mining". pp. 1541–1550. 2002.

[23] K. K. K. M. Ramaboa. "A Study into the Prevalence of Data Mining in South African Companies". Tech. rep., University of Cape Town, Department of Statistical Sciences, 2001.

[24] H. Arndt and H. L. Viktor. "Trend Discovery through Collaborating Data Mining Agents". Tech. rep., University of Pretoria, Department of Informatics, 2000.

[25] M. Matthee and H. L. Viktor. "Data Mining and Organisational Learning: Bridging the Gap". Tech. rep., University of Pretoria, Department of Informatics, 2000.

[26] S. Menard. *Longitudinal Research*. SAGE Publications, Newbury Park, California, 1991.

[27] G. Firebaugh. *Analysing Repeated Surveys*. SAGE Publications, Newbury Park, California, 1997.

[28] F. Davis. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[29] V. Venkatesh and F. A. Davis. "Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies". *Management Science*, vol. 46, no. 2, pp. 186–204, 2000.

[30] E. Rogers. *Diffusion of Innovations*. Free Press, New York, 1995.

[31] V. Venkatesh, M. G. Morris, G. B. Davis and F. D. Davis. "User Acceptance of Information Technology: Towards a Unified View". *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

[32] U. Sekaran (editor). *Research Methods for Business: A Skill Building Approach*. John Wiley and Sons, Ltd, 2nd edn., 1992.

[33] F. Biscombe, M. L. Hart, R. Kraus and S. Naidoo. "An Evaluation of Companies Currently Adopting Data Mining". Tech. rep., University of Cape Town, Department of Information Systems, 2002.