COPYRIGHT AND CITATION CONSIDERATIONS FOR THIS THESIS/ DISSERTATION

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

- NonCommercial — You may not use the material for commercial purposes.

- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

How to cite this thesis
FRAUD DETECTION USING DATA MINING

by

ABEL JACOBUS PIENAAR

SHORT DISSERTATION

submitted in partial fulfillment of the requirements for the degree

MASTER OF COMMERCE

in

COMPUTER AUDITING

in the

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES

at the

RAND AFRIKAANS UNIVERSITY

STUDY LEADER: PROFESSOR ANTON DU TOIT

JOHANNESBURG
NOVEMBER 1997
FRAUD DETECTION USING DATA MINING
DECLARATION

I declare that this research essay submitted to Rand Afrikaans University for the degree Master of Commerce, except to the extent acknowledged in my text, is my own unaided work and has not been submitted previously for any degree to any other university.

ABEL JACOBUS PIENAAR
Main Table of Contents

<table>
<thead>
<tr>
<th>Main Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPSOMMING IN AFRIKAANS</td>
<td>i</td>
</tr>
<tr>
<td>SYNOPSIS</td>
<td>iv</td>
</tr>
<tr>
<td>CHAPTER 1 - INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 - FRAUD DETECTION</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER 3 - DATA MINING</td>
<td>39</td>
</tr>
<tr>
<td>CHAPTER 4 - A MODEL FOR FRAUD DETECTION USING DATA MINING</td>
<td>63</td>
</tr>
<tr>
<td>CHAPTER 5 - SUMMARY OF RESULTS AND CONCLUSION</td>
<td>75</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>78</td>
</tr>
</tbody>
</table>
BEDROGOPSPORING DEUR DIE MYN VAN DATA

deur

ABEL JACOBUS PIENAAR

OPSOMMING VAN SKRIPSIJE
INGEDIEN VIR DIE GRAAD

MAGISTER COMMERCIJ

in

REKENAAROUDITEURING

in die

FAKULTEIT EKONOMIESE EN BESTUURSWETENSKAPPE

aan die

RANDSE AFRIKAANSE UNIVERSITEIT

STUDIE LEIER: PROFESSOR ANTON DU TOIT

JOHANNESBURG
NOVEMBER 1997
Die doel van hierdie opsomming is om die agtergrond, doelwitte, metodiek en gevolgtrekkings van die navorsing weer te gee en is onder die volgende hoofde uiteengesit:

1. Agtergrond en Probleembeskrywing;
2. Doelwitte van die Navorsing;
3. Navorsingsmetodiek en -benadering; en

1. AGTERGROND EN PROBLEEMBESKRYWING

Bedrog is 'n probleem in Suid-Afrika en wêreldwyd wat steeds aan die toeneem is.

Bestuur en ouditeure het nie 'n geïntegreerde model om bedrog in organisasies op te spoor nie en daarom is daar baie bedrog wat onopgespoor bly en nie aangespreek word nie.

Navorsing het bewys dat organisasies tussen twee en vyf persent van hulle omset verloor as gevolg van bedrog en daarom sal 'n geïntegreerde model, bestuur en die ouditeure help om bedrog in 'n organisasie op te spoor.

Dit is dus belangrik vir bestuur en ouditeure om bedrog op te spoor aangesien dit die verskil tussen winsgewendheid en bankrotskap van 'n onderneming kan beteken.
Van die bedrog wat gepleeg word, word gedokumenteer en vasgelê in 'n databasis. Daar is dus baie data wat kan en behoort ontleed te word om bedrog in 'n organisasie op te spoor.

Nuwe tegnologie wat data 'myn', is die afgelope dekade ontwikkel en sal 'n groot hulp vir bestuur en ouditeure wees om bedrog in 'n organisasie op te spoor.

2. DOELWITTE VAN DIE NAVORSING.

Die hoofdoelwit van die navorsing is om vir bestuur, eksterne- en interne ouditeure 'n geïntegreerde model te ontwikkel wat hulle sal help om bedrog in organisasies op te spoor. Die model sluit die proses om data te myn in en dit sal verseker dat die opsporing van bedrog meer doeltreffend en doelmatig sal wees.

Verder moet die geïntegreerde model wat bestuur en ouditeure sal help om bedrog op te spoor, waarde byvoeg tot die organisasie.

3. NAVORSINGSMETODIEK EN BENADERING

Die navorsingsmetodiek wat gevolg is het uit die volgende komponente bestaan:

- 'n Literatuurstudie van bestaande gesaghebbende literatuur oor bedrogopsporing en die myn van data;

- 'n Model van bedrogopsporing is ontwikkel en die verskillende prosesse van kennisontdekking in databasisse (myn van data) ontleed; en
• A integrated model of fraud detection using the mine of data has been developed.

4. GEVOLGTREKKING

The developed integrated fraud detection model will assist and auditors to detect fraud in organisations. Available technology (mine of data) will also help to detect fraud more effectively and efficiently.
SYNOPSIS

Fraud is a major problem in South Africa and the world and organisations lose millions each year to fraud not being detected.

Organisations can deal with the fraud that is known to them, but undetected fraud is a problem. There is a need for management, external- and internal auditors to detect fraud within an organisation. There is a further need for an integrated fraud detection model to assist managers and auditors to detect fraud.

A literature study was done of authoritative textbooks and other literature on fraud detection and data mining, including the Knowledge Discovery Process in databases and a model was developed that will assist the manager and auditor to detect fraud in an organisation by using a technology called data mining which makes the process of fraud detection more efficient and effective.
Chapter 1: INTRODUCTION

Table of Contents

1.1 BACKGROUND AND PROBLEM DESCRIPTION ................................... 2
1.2 OBJECTIVE OF THIS RESEARCH ................................................. 6
1.3 SCOPE AND EXCLUSIONS ......................................................... 7
1.4 DEFINITIONS ............................................................................. 10
1.5 RESEARCH METHODOLOGY ....................................................... 13
1.6 RESEARCH APPROACH ............................................................... 13
1.7 SUMMARY OF RESULTS ............................................................. 14
1.8 CONCLUSION .............................................................................. 16

List of Figures

Figure 1.1 - CALCULATION OF UNDETECTED FRAUD ....................... 5
Figure 1.2 - DEALING WITH FRAUD: A PRO-ACTIVE MODEL .............. 8
Figure 1.3 - STRUCTURE OF SHORT DISSERTATION ......................... 14
1.1 BACKGROUND AND PROBLEM DESCRIPTION

Because of the problem of fraud in business (as discussed below), there is a need for an integrated fraud detection model to assist management and auditors to detect fraud. This integrated model will enable management and auditors of the organisation to identify previously undetected fraud and enable management to investigate and ultimately develop controls to prevent such fraud from recurring in the future. One of the technologies available to help with the detection of fraud is data mining.

In the section that follows the seriousness of the fraud problem and the difficulty to detect it will be discussed. Secondly a method to calculate the potential undetected fraud in an organisation is explained and thirdly the need for using data mining to help to detect fraud in organisations is discussed.

1.1.1 Fraud problem

Fraud is a world wide problem and no organisation is totally immune to this problem.

Commercial crime has increased throughout the world, and South Africa is no exception. It is well documented that the financial costs of commercial crime far exceed the value of all street crime. According to official statistics of reported cases, more than 29 000 cases with an estimated cost of more than R7 billion were under investigation by the SAPS Commercial Branch during 1996. About 80 percent of these cases involved fraud. In addition, the Office for Serious Economic Offences (OSEO) was investigating 33 cases involving approximately R8.5 billion (ISS, 1997:11).
The Nedcor (1997:D19) monthly overview of crime stated that the value of commercial crime cases reported to the commercial police had increased 37 percent to R3,8 Billion last year (1996) compared with 1995, clearly indicating South Africa's vulnerability to organised crime and money laundering.

Commercial crime that includes fraud can be a major threat to the continued existence of an organisation. Albrecht (1995:9) states that thirty percent of all business failures are caused by white-collar crime.

The above mentioned information clearly illustrates the serious threat fraud holds for South African business.

Furthermore fraud is a crime that is difficult to detect. Albrecht (1995:55) states the following: "Fraud is a crime that is seldom seen. If a body is discovered and the person has obviously been murdered, there is no question whether or not a crime has been committed. The dead body can be touched, seen, and even smelled. Likewise, if a bank is robbed, there is no question whether or not a crime has been committed. Everyone in the bank, including customers and employees, witnessed the robbery. In most cases, the entire episode is captured on video and can be replayed for evidence. But with fraud it is not initially certain that a crime has been committed. **Only the fraud symptoms, red flags, or indicators are seen.**"

1.1.2 Calculation of undetected fraud in an organisation

Action on reported (known) cases of fraud in an organisation is not difficult. It is the unreported (unknown) cases that are difficult to respond to.
To begin with, any manager or auditor can calculate the potential fraud problem in an organisation as shown by research outlined below:

Research has shown that, in a broad cross-section of companies in both the US and Europe, an average of between 2 per cent and 5 per cent of turnover is lost to fraud every year. The British home secretary has said that the current cost to industry in the UK is £10 billion annually, whilst losses in the USA have been described as somewhere between what the federal government spends on health and what it spends on defence. For some companies the losses clearly amount to millions. For others it can mean the difference between profit and loss, especially in the areas such as manufacturing where profits as a percentage of turnover are often expressed only in single figures. For hundreds of American banks, it can mean total collapse (Price, 1991:50).

By using the average of between 2 and 5 present of turnover as indicated in the research a person can for example calculate the potential fraud that is undetected in an organisation. If company XYZ Limited's turnover is R100 million per year the company could lose between two and five million rand to fraud. The difference between the reported fraud in the company and the potential fraud is the undetected fraud. For example, if the reported fraud in XYZ Limited is R 1 million, then the undetected fraud will be between R1 million and R 5 million. Refer to Figure 1.1. Calculation of undetected fraud.

After the identification of the potential undetected fraud management should take action to trace the undetected fraud. By developing an integrated fraud detection model management and auditors will be assisted in detecting fraud.
1.1.3 Need for data mining

Fraud is hidden in databases like diamonds in a mine. World-wide the contents and number of databases are growing rapidly. “By 1989 there were an estimated 5 million databases world-wide. The total amount of information in the world is estimated to be doubling every 20 months, and much of this is being stored in computer databases” (Davidson, 1993:28).

Adriaans (1996:2) explains that hundreds of megabytes of data are distributed around the world every day, that the growth is exponential and that more data means less information.

Adriaans (1996:5) also states that the unbridled growth of data will inevitably lead to a situation in which it is increasingly difficult to access the desired information: it will be like looking for a needle in a haystack, only the amount of hay will be growing all the time. It is against this
background that the great interest that is being shown in the new field of ‘data mining’ or KDD (Knowledge Discovery in Databases) is understandable. When mining, enormous quantities of debris have to be removed before diamonds or gold can be found. The analogy that, with a computer, you can automatically find one ‘information-diamond’ among the tons of data-debris in your database, is of course very attractive.

Fayyad (1996:1) states that the explosive growth of many business, government, and scientific databases has far outpaced our ability to interpret and digest this data, creating a need for a new generation of tools and techniques for automated and intelligent database analysis.

These tools and techniques are the subject of the rapidly emerging field of knowledge discovery in databases (KDD) which includes data mining. Data mining technology turns raw data into information that can be used by an organisation and that can be acted upon and can also assist an organisation in detecting fraud and should form part of the integrated fraud detection model.

1.2 OBJECTIVE OF THIS RESEARCH

The objective of this short dissertation is to develop an integrated fraud detection model that includes the use of data mining (included in the knowledge discovery process) to assist management, external auditors and internal auditors to detect fraud.

A further objective is that this fraud detection model should add value to organisations, enabling them to detect fraud in a more effective and efficient manner.
1.3 SCOPE AND EXCLUSIONS

1.3.1 Scope

This short dissertation focuses on fraud detection with specific reference to identifying fraud symptoms, indicators and red flags and the most common frauds committed in business and the use of data mining to detect fraud. The Knowledge Discovery in databases (KDD) process will also be covered.

According to Adriaans (1996:37) this knowledge discovery process (which includes data mining), consists of six stages:

- Data selection;
- Cleaning;
- Enrichment;
- Coding;
- Data Mining; and
- Reporting.

This dissertation will deal with each of the above-mentioned stages.

1.3.2 Exclusions

There are twelve stages for dealing with fraud according to Albrecht (1995:278-281) as indicated in Figure 1.2. This dissertation only deals with stage twelve, namely pro-active fraud auditing (also known as fraud detection).
The stages for dealing with fraud as indicated in Figure 1.2 Albrecht (1995:278-281) are discussed below:

- **Stage 1 - Prevention program**
  Stage 1 is the implementation of prevention programmes designed to educate employees about the seriousness of the problem and to tell them what to do if they suspect fraud.

- **Stage 2 - Incident**
  The second stage will be a fraud incident. Unfortunately, no organisation can totally eliminate fraud, no matter how good its
prevention programmes are. However, detecting fraud early is the key to minimising its impact on the organisation.

- **Stage 3 - Incident reporting**
The third stage involves fraud reporting. It is extremely important that fraud reporting is facilitated.

- **Stage 4 - Investigation**
The procedures are well established and includes: (a) who will conduct the investigation; (b) how the matter will be communicated to management; (c) whether and when law enforcement officials will be contacted; (d) who will determine the scope of investigation; (e) who will determine investigation methods; (f) who will follow up on tips of suspected fraud; (g) who will conduct interviews, review documents, and perform other investigation steps; and (h) who will ultimately determine the corporate response to fraud, disciplines, control, etc.

- **Stage 5 - Action**
A strong prosecution policy must have the support of top management and top managers.

- **Stage 6 - Resolution**
Resolution includes replacing the employee, counselling fellow employees, entering into restitution agreements, and closing the case.

- **Stage 7 - Analysis**
Analysis involves determining how and why the fraud occurred by people in management, audit, security, human resources, control, and finance.
• Stage 8 - Publication
Publication means depersonalising the case and publishing it internally in a newsletter or memo that is distributed to auditors, security personnel and appropriate management and employees.

• Stage 9 - Implementation of controls
This stage involves the explicit implementation of the controls and other measures necessary to prevent future occurrences of fraud.

• Stage 10 - Testing for compliance
This stage includes the testing of controls for compliance.

• Stage 11 - Training
Involves the fraud training of appropriate audit, management, security, and other personnel.

• Stage 12 - Proactive fraud auditing
Pro active fraud auditing (also known as fraud detection) involves proactively auditing for fraud, using discovery sampling and computer query procedures, auditors identify fraud characteristics and search databases looking for them. This short dissertation covers only Stage 12 of the pro-active model for dealing with fraud.

1.4 DEFINITIONS

A definition of the following words is needed to clarify the use of the words in this short dissertation:
• Data

Adriaans (1996:133) defines data as any symbolic representation of facts or ideas from which information can potentially be extracted.

Quantities or characters operated on by a computer (Oxford, 1990:294).

• Database

A database is a large collection of data, mostly stored in a computer system (Adriaans, 1996:133).

A structured set of data held in a computer (Oxford, 1990:294).

• Data Mining

The actual discovery phase of a knowledge discovery process (also refer to KDD) (Adriaans, 1996:133).

"Data mining is the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques. The Gartner Group" (Berry, 1997:96).

• Data warehouse

Adriaans (1996:133) defines a data warehouse as a subject-orientated, integrated, time-variant, non-volatile collection of data in support of management.
• Effective

Having a definite or desired effect (Oxford, 1990:374).

• Efficient

Productive with minimum waste or effort (Oxford, 1990:374).

• Fraud

Snyman (1989:504) defines fraud as the unlawful and intentional making of a misrepresentation which causes actual prejudice or which is potentially prejudicial to another.

Fraud is criminal deception; the use of false representations to gain an unjust advantage (Oxford, 1990:467).

• Fraud Detection

Detection of fraud includes the steps or actions taken to discover that fraud has been committed. Detection does not include investigative procedures taken to determine motive, extent, method of embezzlement, or other elements of the dishonest act (Albrecht, 1995:53).

• KDD

KDD is knowledge discovery in databases, the non-trivial extraction of implicit, previously unknown, and potentially useful information from data (Adriaans, 1995:135).
1.5 RESEARCH METHODOLOGY

The following methodology was employed in performing the research:

- A literature survey has been done of existing authoritative textbooks and other literature, as well as discussions with people with technical knowledge, on fraud detection and data mining (chapters 2 and 3).

- With all the information obtained in the literature survey, an integrated model for fraud detection using data mining, has been developed (chapter 4).

1.6 RESEARCH APPROACH

This dissertation is summarised in Figure 1.3 (p.14) and the rationale is explained below:

- Chapter 1 discusses the problem of fraud and establishes the need for an integrated fraud detection model.

- In chapter 2, the role of management, external and internal auditors in fraud detection is discussed. The fraud detection process is also discussed in detail with specific reference to fraud symptoms, red flags, fraud indicators and the most common types of fraud.

- Chapter 3 gives a description of data mining and discusses the evolution and techniques of data mining. The process of knowledge discovery, of which data mining is a part, is also described and reference is made to different frameworks.
• In chapter 4, Fraud detection and data mining processes (discussed in chapter 2 and 3) are integrated and a model for fraud detection using data mining, is developed.

• Finally, chapter 5 summarises the results, and concludes the short dissertation and indicates whether the objectives of this short dissertation have been met.

Figure 1.3 - Structure of short dissertation (Prepared by author)

1.7 SUMMARY OF RESULTS

Chapter 2 - Fraud Detection
The role of management, external and internal audit are discussed in this chapter. They all have a very important role to play in detecting fraud as could be seen from their responsibilities.
A fraud detection process was also developed and the steps to detect fraud were set out as follows:

- Step 1 - Identify the fraud symptoms, indicators and red flags in the organisation;
- Step 2 - Identify the common fraud areas in the organisation;
- Step 3 - Identify the high fraud risk areas; and
- Step 4 - Audit for fraud by using a technology called data mining.

Chapter 3 - Data Mining

Data mining is part of the knowledge discovery process and cannot be viewed in isolation. The knowledge discovery process was discussed in detail in this chapter and the main steps in the knowledge discovery process are the following:

- Step 1 - Information requirement;
- Step 2 - Data Selection;
- Step 3 - Data Cleaning;
- Step 4 - Data Enrichment;
- Step 5 - Data Coding / Reduction;
- Step 6 - Data Mining;
- Step 7 - Interpretation / Reporting;
- Step 8 - Action; and
- Step 9 - Measurement and Evaluation.

The fraud detection process as discussed in chapter 2 resorts under Step 1 of the knowledge discovery process and takes place before the real knowledge discovery.

In chapter 4 the knowledge discovery processes discussed in this chapter are combined to best fit the integrated fraud detection model.
Chapter 4 - Model for Fraud detection using Data Mining

An integrated model for fraud detection using data mining was developed which will assist management and the auditor to understand fraud detection and to detect fraud using data mining.

Management and the auditor need to understand that detecting fraud is not an easy task, but if the above-mentioned model is followed, it would help them to detect fraud in an organisation.

1.8 CONCLUSION

Fraud is a serious problem for business in South Africa and the world. Management and auditors need a model to detect fraud in organisations. An effective and efficient way of detecting fraud is by using a technology called data mining.

The objectives set out in this chapter have been reached: An integrated fraud detection model has been developed for detection of fraud by using the data mining technology which forms part of the knowledge discovery process and the model will assist management, external and internal auditors to detect fraud when it is required of them to do so. This model will also add value to organisations and enable them to detect fraud in a more effective and efficient way.

A suggested area for further research could be to explore the other stages of the process of dealing with fraud as described in paragraph 1.3.2 (p.8) for example prevention, and link them to a relevant technology that can assist management and the auditor in their fight against fraud.
## Chapter 2: FRAUD DETECTION

### Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 INTRODUCTION</td>
<td>18</td>
</tr>
<tr>
<td>2.2 NATURE OF LITERATURE SURVEY</td>
<td>18</td>
</tr>
<tr>
<td>2.3 SCOPE, LIMITATIONS AND EXCLUSIONS</td>
<td>19</td>
</tr>
<tr>
<td>2.4 OBJECTIVES OF THIS CHAPTER</td>
<td>20</td>
</tr>
<tr>
<td>2.5 ROLE OF MANAGEMENT AND AUDITORS IN FRAUD DETECTION</td>
<td>21</td>
</tr>
<tr>
<td>2.6 FRAUD SYMPTOMS, RED FLAGS AND FRAUD INDICATORS</td>
<td>24</td>
</tr>
<tr>
<td>2.7 COMMON FRAUD AREAS</td>
<td>35</td>
</tr>
<tr>
<td>2.8 FRAUD DETECTION PROCESS</td>
<td>36</td>
</tr>
<tr>
<td>2.9 CONCLUSION</td>
<td>38</td>
</tr>
</tbody>
</table>

### List of Figure

- Figure 2.1 - FRAUD DETECTION PROCESS                                   36
2.1 INTRODUCTION

Fraud is a serious business problem in South Africa today. Management and auditors need to know the process of detecting fraud in organisations to assist them in detecting fraud.

By detecting and dealing with fraud in an organisation, the organisation will be more profitable. The detection of fraud will also help the organisation to identify the areas of vulnerability. The organisation can then create and implement the necessary controls in those vulnerable areas to prevent fraud in the future.

2.2 NATURE OF LITERATURE SURVEY

To ensure credibility and acceptance of the findings and proposals of this short dissertation, it is essential that the underlying concepts should be based on authoritative views and be generally accepted among management in general and audit professionals. References have been restricted to those published by fraud experts and to authoritative technical books and publications by computer experts. The main authors/publishers are:

- Albrecht, Wernz and Williams;
- Huntington and Davies;
- Adriaans and Zantinge; and
- Berry and Linoff.

The reasons for choosing these authors are the following:

- W. Steve Albrecht is the Director of the J. Willard and Alice S. Marriott School of Management at Brigham Young University. He has consulted with numerous organisations on the problem of fraud, including IBM,
the FBI, Deere & Co, Caterpillar, Alcoa, Bank of America, and many others. Albrecht is the past president of the National Association of Certified Fraud Examiners.

- Ian Huntington is Head of Fraud Investigation at KPMG Forensic Accounting, London. He is one of KPMG's fraud specialists. KPMG Forensic Accounting has 90 specialist partners and staff in the UK assigned to investigate fraud and assisting with the litigation process full time.

- Pieter Adriaans is a director of Syllogic, where he is responsible for the development of tools for the management of client/server systems and databases. The basis for Syllogic's activities in this area is the integration of artificial intelligence techniques, machine learning, object orientation and database management. This author has published extensively in the area of systems management, machine learning and client/server.

- Michael J.A. Berry is a principal at MRJ Technology Solutions, a Pennsylvania-based firm specialising in data warehousing and data mining for business applications. He also conducts seminars on data mining for Fortune 100 companies.

### 2.3 SCOPE, LIMITATIONS AND EXCLUSIONS

#### 2.3.1 Scope

Relevant literature was surveyed for any reference to fraud detection and data mining. The literature was then further surveyed to understand the processes involved in data mining (KDD) and the detection of fraud to develop an integrated model for fraud detection using data mining.
2.3.2 Limitations and exclusions

The following limitations and exclusions were placed on the scope of the literature survey:

**Fraud Detection**

- There is limited literature available that deals specifically with fraud detection.
- Available literature is either criminal law and legal literature or literature covering the investigation of fraud.
- Only fraud detection was focussed on, the other stages for dealing with fraud is excluded (refer to paragraph 1.3.2).
- It was possible to get authoritative literature on fraud detection as used in this short dissertation.

**Data Mining**

- There is limited literature available on data mining because it is a new field of technology.
- Available literature was directed more towards using data mining for marketing, sales and customer support.
- It was possible to get authoritative literature on data mining as used in this short dissertation.
- Data mining software references are limited to Netmap.

2.4 OBJECTIVES OF THIS CHAPTER

The primary objective of this chapter is to describe the role of management, external and internal auditors in detecting fraud, as well as a utilising a method to detect fraud in a business environment by identifying
fraud symptoms, red flags and fraud indicators and the more important common fraud areas in business to assist management and the auditor to detect fraud.

The secondary objective is to develop a fraud detection process diagramme for building an integrated fraud detection model in chapter 4, that will help management, external and internal auditors to detect fraud.

2.5 ROLE OF MANAGEMENT AND AUDITORS IN FRAUD DETECTION

It is important to understand the role of management, external audit and internal audit in the detection of fraud. They all have a role to play in the detection of fraud as discussed below.

2.5.1 Role of Management and External Auditors

In the statement of SAICA (1992: AU 005:3) the responsibility of management is described as follows:

"The responsibility for the prevention and detection of illegal acts, other irregularities and errors within the entity rests with management. Management of the entity act in a fiduciary capacity in relation to the property that is under their control and it is their responsibility to ensure that the entity's operations are conducted in accordance with all relevant legal obligations."

In the statement of SAICA (1992: AU 005:4) the responsibility of the external auditor is described as follows:

"The auditor is not responsible for preventing illegal acts, other irregularities and errors. The fact that an annual audit is carried out and
that the auditor would normally report on illegal acts, other irregularities and errors which come to his/her attention may act as a deterrent. Reports to management may highlight areas where illegal acts, other irregularities or errors could occur and suggest practical steps which, if implemented, may help to prevent illegal acts, irregularities or errors.

The auditor's responsibility is to plan, perform and evaluate audit work in such a way that there is a reasonable expectation of detecting material misstatements in the financial information which result from illegal acts, other irregularities and errors, and which affect the fair presentation of the financial information."

Sawyer (1988:1004), refers to the case of In re Kingston Cotton Mill Co., Limited (No. 2), (1896 2 Ch.279) where the court determined the following: "It is the duty of an auditor to bring to bear on the work he has to perform that skill, care and caution which a reasonably competent, careful, cautious auditor would use... An auditor is not bound to be a detective, or, as was said to approach his work with suspicion or with the foregone conclusion that there is something wrong. He is a watchdog, but not a bloodhound ... If there is anything calculated to excite suspicion he should probe it to the bottom, but in the absence of anything of that kind, he is only bound to be reasonably cautious and careful..."

Taking note of the above-mentioned, it is in my opinion still important that the external auditor will identify the material frauds during the external audit process. To assist him with this he should know the fraud symptoms, red flags, fraud indicators and common fraud areas.

Grant Thornton (1997:1) says that a new development is that the Auditing Standards Board (ASB) of the American Institute of Certified Public Accountants (AICPA) has released its proposed Statement on Auditing
Standards (SAS) titled Consideration of fraud in a Financial Statement Audit in May 1996 and that the driving force for the new standard is to enhance auditors' performance in detecting material misstatements of financial statements because of fraud. A key feature of the proposed standard is a list of fraud factors that auditors should look for in every engagement.

These fraud factors relate to fraud symptoms, indicators and red flags which are discussed in section 2.6 of this chapter.

2.5.2 Role of Internal Auditors

Sawyer (1988:1006) refers to the Statement on Internal Auditing No. 3 which expands on the responsibilities of internal auditors with respect to fraud. The major conclusions of the statement regarding the detection of fraud are as follows:

"Internal auditors should have sufficient knowledge of fraud to be able to identify indicators that fraud might have been committed. If significant control weaknesses are detected, additional tests conducted by internal auditors should include tests directed toward identification of other indicators of fraud. Internal auditors are not expected to have equivalent to that of a person whose primary responsibility is to detect fraud."

Fraud is therefore a significant business issue and the internal auditors should be trained in and know fraud symptoms, red flags, fraud indicators and common fraud areas to help them detect fraud.
2.6 FRAUD SYMPTOMS, RED FLAGS AND FRAUD INDICATORS

The detection of fraud begins with identifying symptoms, indicators, or red flags that can be associated with fraud (Albrecht, 1995:53).

The first step in detecting fraud in a business is to identify the fraud symptoms, indicators, or red flags in the business and it is very important for management and auditors to know what the fraud symptoms, indicators or red flags are to detect or identify a possible case of fraud. For this reason fraud symptoms, indicators or red flags are discussed in detail in this section.

Fraud symptoms, indicators and red flags are usually the only elements of an undetected fraud that are visible. The more symptoms, indicators and red flags recognised in an organisation, the greater the probability of fraud.

Fraud symptoms, indicators or red flags will be discussed under the following headings:

- Operating performance anomalies;
- Organisational structure;
- Management characteristics;
- Accounting anomalies;
- Relationships with other parties;
- Extravagant lifestyle;
- Unusual behaviour;
- Tips and complaints; and
- Investment fraud symptoms.
2.6.1 Operating performance anomalies

Albrecht (1995:60) states that operating performance anomalies include red flags identified by analysing financial statements and other organisation reports. Common operating performance anomalies include:

- Unexplained changes in financial statement balances;
- Operating on a crisis basis;
- Urgent need to report favourable earnings;
- Unusual or large and profitable transactions near the end of accounting periods;
- Deteriorating quality of earnings;
- Insufficient capital;
- High debt or interest burdens;
- Difficulty in collecting receivables, or other cash flow problems;
- Expenses increasing faster than revenues;
- Dependence on only one or two products; and
- Significant litigation.

Albrecht (1995:60) gives the following true example to explain the operating performance anomaly:

"An example in which some of these symptoms indicated fraud was the recent financial statement fraud discovered at Comptronix Corporation. The fraud involved misstated financial statements and was supposedly recognised by a college professor who, after analysing the company's annual report, noticed that sales increased but receivables didn't and that inventory increased but accounts payable and purchases didn't. These unexplained and irrational changes in financial statement relationships did not make sense. They occurred because the management of Comptronix overstated sales and inventory to increase reported net income."
Other operating performance anomalies (business risks) are:

- Profits well in excess of industry norm; and

### 2.6.2 Organisational structure anomalies

Organisational structure symptoms include attributes of an organisation that are unreasonable and/or do not have a legitimate reason for existence. Common symptoms of this type include:

- Unduly complex business structure;
- Lack of effective internal audit staff;
- High-risk industry;
- Severe obsolescence of assets;
- Changes in executives or directors; and

An specific example regarding organisational structure anomalies given by Albrecht (1995:63) was that of Charles Keating, creating a complex business structure at Lincoln Savings involving over 15 separate business entities. Having various subsidiaries sell and buy real estate, while other subsidiaries borrowed and loaned money, the impression was created that transactions were separate from each other. This complex business structure was set up to hide management fraud.

### 2.6.3 Management characteristics

Management characteristic symptoms include attributes of executives that signal a strong motivation to be dishonest or a high risk of fraud. According to Albrecht (1995:62) common fraud indicators of this type are:

- Executives with high personal debt or financial needs;
- Executives with questionable or criminal backgrounds;
• Executives who are involved extensively in gambling or speculative transactions;
• Dishonest or unethical management;
• Executives whose financial success is closely tied to the success of the organisation; and
• Executives who don't have lives separate from the organisation.

Huntington (1994:9-10) identifies other management characteristic symptoms as:
• Autocratic management style; and
• Mismatch between personality and status.

The following symptoms could be added to the list:
• Executives that disregard internal controls and;
• Executives that disregard internal and external audit reports.

2.6.4 Accounting anomalies

Albrecht (1995:62) states that common fraud symptoms involving source documents (such as checks, sales invoices, purchase orders, purchase requisitions) and receiving reports includes the following:
• Missing documents;
• Stale items on bank reconciliations;
• Excessive voids or credits;
• Common names or addresses of payees or customers;
• Increased past due accounts;
• Increased reconciliation items;
• Alterations on documents;
• Duplicate payments;
• Second endorsements on checks;
- Document sequences that don't make sense;
- Questionable handwriting on documents; and
- Photocopied documents.

"A major difference between auditors and examiners who detect fraud and those who don't, is that most auditors merely match documents to see whether support exists and is adequate. Auditors and examiners who detect fraud go beyond ascertaining the mere existence of documents to determine whether the documents are real or fraudulent, whether the expenditures make sense, and whether all aspects of the documentation are in order" (Albrecht, 1995:57).

Business Fraud Detection Services (1997:2) says that when funds, merchandise, or assets are stolen and not covered by a fictitious entry, the general ledger will be out of balance. An inventory of the merchandise or cash is needed to confirm the existence of the missing assets.

It is important to be aware of accounting anomalies, it could be an easy way to detect fraud.

2.6.5 Internal control weaknesses

Internal control includes the control environment, the accounting system and control procedures. Albrecht (1995:93) states that common internal control fraud symptoms are:
- Lack of segregation of duties;
- Lack of physical safeguards;
- Lack of independent checks;
- Lack of proper authorisation;
- Lack of proper documents and records;
- Overriding of existing controls;
Inadequate accounting system and that many studies have found that the element most common in frauds is the overriding of existing internal controls.

Usually internal controls do exist in an organisation but are overridden by management and staff to commit a fraud.

2.6.6 Analytical anomalies

Albrecht (1995:98-99) explains that analytical fraud symptoms are procedures or relationships that are too unusual or too unrealistic to be believable. Common examples of analytical symptoms include:

- Unexplained inventory shortages or adjustments;
- Deviations from specifications;
- Increased scrap;
- Excess purchases;
- Too many debit or credit memos;
- Significant increases or decreases in account balances;
- Physical abnormalities;
- Cash shortages or overages;
- Excessive late charges;
- Unreasonable expenses or reimbursements; and
- Strange financial statement relationships such as:
  - Increased revenues with decreased inventory;
  - Increased revenues with decreased receivables;
  - Increased revenues with decreased cash flows;
  - Increased inventory with decreased payables;
  - Increased volume with increased cost per unit; and
  - Increased volume with decreased scrap.

Analytical tests can assist management and the auditor to detect fraud.
2.6.7 Relationships with other parties

Albrecht (1995:104) states that an organisation cannot exist without interacting extensively with other organisations, government agencies and individuals. Specific fraud symptoms involving relationships with others are:

- Relationships with lawyers:
  - Significant litigation; and
  - Frequent changes in legal counsel.

- Relationships with auditors:
  - Frequent changes in outside auditors;
  - Reluctance to give auditors needed data;
  - Lack of internal auditors; and
  - Adverse opinion or disclaimer of opinion on financial statements.

- Relationships with board members:
  - Unexplained resignation of board members; and
  - Frequent turnover of board members.

- Relationship with regulators and IRS (e.g. South African Revenue Services):
  - Revoke licences;
  - Frequent review by regulators;
  - Identification by regulators as being high-risk;
  - Significant or frequent tax adjustments; and
  - Continuing problems with the IRS.

- Relationships with management:
  - High turnover of management; and
  - New management.

- Relationships with banks and other lenders:
  - High debt;
  - Violation of debt restrictions;
  - Use of several different banks;
Inability to borrow, or poor credit; and
Existence of large contingent liabilities.

- Relationship with related parties:
  Existence of related-party transactions.

- Relationships with other companies:
  Pressure to merge, sell or be taken over; and
  Restructuring.

- Relationships with vendors or customers:
  High-volume new customers or vendors.

2.6.8 Extravagant lifestyle

Albrecht (1995:128) states that most people who commit fraud are under financial pressure. Sometimes the pressures are real, and sometimes mere greed. Once the perpetrators meet their financial needs, they usually continue to steal, using the embezzled funds to improve their lifestyles. Often they buy new cars. They may buy other expensive toys, take vacations, remodel their homes or move into more expensive houses, buy expensive jewellery or clothes, or just start spending more money on food and other day to day living.

"Lifestyle changes are often the easiest of all symptoms to detect. If managers, co-workers, and others would pay attention, they would notice embezzlers living lifestyles their incomes don't support" (Albrecht, 1995:130).

Examples of expensive lifestyles that were given by Huntington (1994:11) were those of a manager who used to buy his staff champagne every Friday afternoon and another manager who changed his Jaguar every year.
Crumbley (1997:4) says that according to Joseph T. Wells, the founder of the Association of Fraud Examiners, the simple recipe for spotting fraud perpetrators is to look for flashy clothes and jewellery.

A fraud symptom would for example be if a low level clerk came to work in an expensive motor vehicle for e.g. a Ferrari!

2.6.9 Unusual behaviour

Albrecht (1995:125) says that research in psychology reveals that when a person commits a crime, he or she becomes engulfed by emotions of fear and guilt. These emotions express themselves in an extremely unpleasant sensation called stress. Some recognisable behavioural changes are:

- Insomnia;
- Increased drinking;
- Taking of drugs;
- Unusual irritability and suspiciousness;
- Inability to relax;
- Lack of pleasure in things the person usually enjoys;
- Fear of getting caught;
- Inability to look people in the eye;
- Showing embarrassment around friends, co-workers and family;
- Defensiveness or argumentativeness;
- Unusual belligerence in stating opinions; and
- Confessing.

Albrecht (1995:125) states that it is not any particular behaviour that signals fraud, rather changes in behaviour.
Business Fraud Detection Services (1997:1) explains that the perpetrator will often display unusual behaviour which, when taken as a whole, is a strong indicator of fraud. The fraudster may not ever take a vacation or call in sick in fear of being caught. He or she may not delegate or assign out work even when overloaded.

2.6.10 Tips and complaints

Auditors are often criticised for not detecting more fraud. Albrecht (1995:118) states that because of the nature of fraud, auditors are often in the worst position to detect its occurrence. He goes further by saying that one large company, that uncovered over 1,500 individual cases of fraud during 1992, discovered 43 percent of these cases as a result of customer complaints or employee tips. Whenever tips or complaints are received, they must be treated with care and considered only as fraud symptoms. Individuals should always be considered innocent until proved guilty and should not be unjustly suspected or indicated.

Business Fraud Detection Services (1997:1) explains that complaints have been known to be some of the best sources of fraud detection and should be taken seriously. Although, all too often the motives for the complaint may be suspect, the allegation usually has merit that warrants further investigation.

2.6.11 Investment fraud symptoms

Albrecht (1995:64) says that with investment fraud, perpetrators usually make fraudulent promises or misstatements of fact to induce people to make investments. There are numerous red flags or fraud symptoms that signal potential investment fraud:

- Unreasonable promised rates of return;
• Investments that do not make sound business sense;
• Pressure to get in early on the investment;
• Use of special tax loopholes or tax avoidance schemes;
• A business that is new in town and does not offer an adequate history of its principals and their operations in previous locations;
• A business with history of bankruptcy or scandals;
• Appraisal figures and/or financial claims that have not been soundly verified;
• Project dependency on kickbacks, complicated marketing schemes, special concessions to people who have money, or unwritten deals that can’t be talked about because of domestic or foreign laws;
• Unaudited financial reports or adverse opinions given on financial reports;
• Investments that assume continued inflation or appreciation in predicting attractive rates of return and that are unrealistic over time;
• Investment success that is dependent on someone’s “unique expertise” for financial success;
• Representation of the emotional desirability of holding an investment as its principal attraction;
• Insufficient verification or guarantee of an investment;
• Dependency on high financial leverage for success;
• Investor liability for debts not paid;
• Luxurious lifestyles of principals, even though the business is relatively new;
• An investment that is not suitable for your risk tolerance;
• Pressure to put all your savings into a particular investment;
• Inability to withdraw or liquidate the investment; and
• Inducements that make investors feel sorry for the principals and/or put in additional money to help overcome temporary problems.
2.7 COMMON FRAUD AREAS

It is also very important for management and the auditor to know the common fraud areas. This will help them to identify the high risk fraud areas in an organisation.

If the manager and the auditor know the areas in the organisation that is more susceptible to fraud and also know the symptoms, indicators or red flags as discussed above, he or she would be in a better position to detect fraud than the manager or auditor who do not know the common fraud areas, symptoms, indicators or red flags.

In the following section common forms of fraud in a business are discussed. These common forms of fraud are not intended to be a complete list of all types of fraud possible, but rather a list that will sensitise management and the auditors to the fact that fraud do exist and that they should take note of it.

In Sawyer (1988:1016) forty common forms of fraud are discussed. The more important forms are:

- Failing to record sales of merchandise, and pocketing the cash;
- Overloading expense accounts or diverting advances for personal use;
- Lapping collections on customers' accounts;
- Pocketing payments on customers' accounts, issuing receipts on scraps of paper or in self-designed receipt books;
- Issuing credit for false customer claims and returns;
- Failing to make bank deposits daily or depositing only part of the money;
- Altering dates on deposit slips to cover stealing;
- Carrying fictitious extra help on payrolls or increasing rates or hours;
- Carrying employees on the payroll beyond actual severance dates;
• Using personal expenditure receipts to support false paid-out items;
• Increasing amounts of suppliers' invoices through collusion;
• Billing stolen merchandise to fictitious accounts;
• Falsifying inventories to cover thefts or delinquencies;
• Charging personal purchases to the company through misuse of purchase orders;
• Obtaining blank cheques (unprotected) and forging the signature; and
• Permitting special prices or privileges to customers or granting business to favoured suppliers for “kickbacks”.

2.8 FRAUD DETECTION PROCESS

As could be seen from the above section it is important for management and the auditor to identify the fraud symptoms, indicators and red flags as well as the common fraud areas to assist them to detect fraud. Thereafter, management and the auditor can identify the high fraud risk areas to 'audit' for fraud and thus detecting it. A fraud detection process diagramme (Figure 2.1) deducted from the above would be:

Figure 2.1 - Fraud detection process (Prepared by author)

<table>
<thead>
<tr>
<th>Identify the fraud symptoms, Indicators, or red flags</th>
<th>Identify the common fraud areas In the organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Identify the high fraud risk areas in the organisation</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>Audit for fraud in the high fraud risk areas in the organisation by using data mining (Knowledge discovery process).</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.1 shows the relationship between the different steps that should be taken to detect fraud:

(1) **Identify the fraud symptoms, indicators and red flags**

This is the first step in identifying an undetected fraud. Management and auditors should know the fraud symptoms, indicators and red flags to assist them to identify fraud. The higher the number of fraud symptoms, indicators and red flags present in an organisation, the more probable fraud is in that organisation.

(2) **Identify the common fraud areas in the organisation**

This step could be done simultaneously with the first step, which is why it is shown on the same level as the first step. Management and auditors should know the common areas where fraud could occur in their organisation. This would help them to identify the high fraud risk areas.

(3) **Identify the high fraud risk areas in the organisation**

After identifying the common fraud areas and the fraud symptoms, indicators and red flags in the organisation, management and the auditor should be able to identify the high fraud risk areas in the organisation. The high fraud risk areas are those areas where fraud is more likely to occur in an organisation. Management and auditors can then draw up a document indicating the high fraud risk areas called a Business Fraud Risk Profile.
(4) Audit for fraud in the high fraud risk areas of an organisation

After management or the auditor have identified the areas where the organisation is more likely to have fraud, they can now audit for fraud in that area. To audit for fraud without the help of technology would be ineffective and inefficient. A technology that can assist management and the auditor to detect fraud is data mining (discussed in chapter 3).

2.9 CONCLUSION

The role of management, external and internal audit has been discussed in this chapter. They all have a very important role to play in detecting fraud as could be seen from their responsibilities discussed above.

A fraud detection process was also developed and the steps to detect fraud described. They are:

- Step 1 - Identify the fraud symptoms, indicators and red flags in the organisation;
- Step 2 - Identify the common fraud areas in the organisation;
- Step 3 - Identify the high fraud risk areas; and
- Step 4 - Audit for fraud by using a technology called data mining (Knowledge Discovery Process - discussed in chapter 3).
Chapter 3: DATA MINING

Table of Contents

3.1 INTRODUCTION.......................................................... 40
3.2 OBJECTIVES OF THIS CHAPTER........................................... 40
3.3 DATA MINING............................................................... 40
3.4 THE DATA MINING PROCESS - VIRTUOUS CYCLE
    - BERRY FRAMEWORK................................................... 50
3.5 THE KNOWLEDGE DISCOVERY PROCESS
    - ADRIAANS FRAMEWORK................................................. 53
3.6 THE KNOWLEDGE DISCOVERY PROCESS
    - BRACHMAN AND ANAND FRAMEWORK................................ 58
3.7 CONCLUSION.............................................................. 62

List of Figures and a Table

Figure 3.1 - SERIOUS FRAUD OFFICE - NETMAP GRAPH.............. 49

Figure 3.2 - THE FOUR STAGES OF THE VIRTUOUS CYCLE OF
    DATA MINING.......................................................... 50

Figure 3.3 - THE KDD PROCESS - ADRIAANS FRAMEWORK............... 54

Figure 3.4 - THE KDD PROCESS - BRACHMAN & ANAND.................. 59

Table 3.1 - STEPS IN THE EVOLUTION OF DATA MINING............... 43
3.1 INTRODUCTION

Data mining is a new technology and has played a more predominant role in discovering treasures of information in databases in recent years. Data mining is part of the Knowledge Discovery Process as discussed in this chapter.

CWI (1997:1) states that companies accumulate more information in databases each year. As a result, these databases contain treasures of information on many of the company's processes. This information can be used to improve such processes and allows the company to detect trends and react to them. However, this information is hidden in the mountains of data and cannot be discovered using conventional database management systems. The solution is data mining.

3.2 OBJECTIVES OF THIS CHAPTER

The objective of this chapter is to describe what data mining is, the evolution of data mining and techniques used in data mining.

A further objective is to describe the more important knowledge discovery processes and to determine the relationship between data mining and fraud detection.

From the Knowledge Discovery Process frameworks discussed in this chapter, a model will be developed for the integrated fraud detection model described in chapter 4.
3.3 DATA MINING

3.3.1 Introduction to data mining

Here are some of the more important ideas on data mining:

- "Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. The Gartner Group" (Clearsoft, 1997).

- "Data mining provides a way to analyse organisational data, discovering hidden patterns within tremendous quantities of data, and providing an organisation with information it may not have been able to discover" (Toyota, 1997a).

- "Data mining is defined as the extraction of hidden predictive and descriptive information from large databases" (Toyota, 1997b).

- "The goal of data mining is still much like that of artificial intelligence. The data mining user doesn't exactly pose a question as much as ask the system to use a model to discover past patterns that predict future behaviour. This can result in valuable, previously unknown facts" (Gerber, 1996:41).

- Adriaans (1996:5) states that there is confusion about the exact meaning of the terms 'data mining' and 'KDD' as many authors regard them as synonymous. At the first international KDD conference in Montreal in 1995, it was proposed that the term KDD be employed to describe the whole process of extraction of knowledge from data. In this context, knowledge means relationships and patterns between
data elements. It was further proposed that the term 'data mining' should be used exclusively for the discovery stage of the KDD process. A more or less official definition of KDD is: 'the non-trivial extraction of implicit, previously unknown and potentially useful knowledge from data.' So the knowledge must be new, not obvious, and one must be able to use it.

It is clear that data mining is technology that can assist management and the auditor to detect the hidden treasures in their data. The hidden treasures referred to here are the undetected cases of fraud. Data mining is a part of the overall Knowledge Discovery Process in databases.

3.3.2 Evolution of data mining

Pilot Software (1997a) says that data mining takes the evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now mature, namely:

- Massive data collection;
- Powerful multiprocessor computers; and
- Data mining algorithms.

Data mining algorithms embody techniques that have existed for at least 10 years but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods. In the evolution from business data to business information, each new step has built upon the previous one. For example, dynamic data access is critical for drill-through in data applications and the ability to store large databases is critical to data mining. From the user's point of
view, the four steps listed in Table 3.1 were revolutionary because they allowed new business questions to be answered accurately and quickly.

Pilot Software (1997a) further states that the core components of data mining technology have been under development for decades in research areas such as statistics, artificial intelligence, and machine learning. Today the maturity of these techniques, coupled with high-performance relational database engines and broad data integration efforts, make these technologies practical for current data-warehousing environments.

Table 3.1 - Steps in the evolution of data mining (Pilot Software, 1997)

<table>
<thead>
<tr>
<th>Evolutionary Step</th>
<th>Business Question</th>
<th>Enabling Technologies</th>
<th>Product Providers</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection (1960's)</td>
<td>“What was my total revenue in the last five years?”</td>
<td>Computers, tapes, disks</td>
<td>IBM, CDC</td>
<td>Retrospective, static data delivery</td>
</tr>
<tr>
<td>Data access (1980's)</td>
<td>“What were unit sales in New England last March?”</td>
<td>Relational databases (RDBMS), Structured Query language (SQL), ODBC</td>
<td>Oracle, Sybase, Informix, IBM, Microsoft</td>
<td>Retrospective, dynamic data delivery at record level</td>
</tr>
<tr>
<td>Data Warehousing &amp; Decision Support (1990's)</td>
<td>“What were unit sales in New England last March? Drill down to Boston.”</td>
<td>Online analytic processing (OLAP), Multidimensional databases, data warehouses</td>
<td>Pilot, Comshare, Arbor, Cognos Microstrategy</td>
<td>Retrospective, dynamic data delivery at record level</td>
</tr>
<tr>
<td>Data Mining (Emerging today)</td>
<td>“What is likely to happen to Boston unit sales next month. Why?”</td>
<td>Advanced algorithms, Multiprocessor computers, massive databases.</td>
<td>Pilot, Lockheed, IBM, SGI, Numerous startups (nascent industry)</td>
<td>Prospective, proactive information delivery</td>
</tr>
</tbody>
</table>
The evolution of data mining (Table 3.1, p.43) indicate that data mining is a new technology that has evolved as software and hardware industries grew.

### 3.3.3 Data mining techniques

The more important data mining techniques are the following:

- **Market Basket Analysis**

  Market basket analysis is a form of clustering used for finding groups of items that tend to occur together in a transaction (or market basket). The models that it builds give the likelihood of different products being purchased together and can be expressed as rules. It is closely related to analyses in the retail industry where information on products purchased together may be the only data available for mining customer patterns (Berry, 1997:119).

- **Memory-Based Reasoning (MBR)**

  MBR is a directed data mining technique that uses known instances as a model to make predictions about unknown instances. MBR looks for the nearest neighbours in the known instances and combines their values to assign classification or prediction values. For instance, we might maintain a database of claims and whether they were adjusted after investigation. If we want to determine if a new claim warrants further investigation, we would find similar claims - neighbours - in the database and make the 'investigate-further' or 'pay-immediately' decision based on the status of the neighbours. The distance to the neighbours gives a measure of correctness of the results (Berry, 1997:120).
• Cluster Detection

Cluster detection is the building of models that find data records that are similar to each other. These clumps of self-similarity are called clusters. This is inherently undirected data mining, since the goal is to find previously unknown similarities in the data. There are several techniques for finding clusters including geometric methods, statistical methods, and neural networks (Berry, 1997:121).

• Link Analysis

Link analysis follows relationships between records to develop models based on patterns in the relationships. This is an application of graph theory constructs to data mining. Relationships between customers are becoming increasingly important, especially as marketing groups focus more on customers, households and economic marketing units instead of specific accounts. The largest area where it is applied is actually in the law enforcement area, where clues about crimes are linked together to help solve them. The few tools available focus on visualising the links, rather than analysing the patterns (Berry, 1997:121).

KPMG (1997:3) explains relationship mapping as follows:
"Relationship mapping is relatively new and uses a powerful mini-computer to map relationships from transaction files. The powerful feature of this type of program is that the output it generates is graphical i.e. the software uses certain rules and algorithms to draw relationships between items in a graphical format. This make the investigators work easier in the sense that instead of sifting through thousands of transactions, the computer can do this quicker and more effectively and highlight the areas where the person has to concentrate his/her efforts."
• Decisions Trees and Rule Induction

Decision trees are powerful models produced by a class of techniques that include classification and regression trees (CART) and chi-squared automatic induction (CHAID). Decision trees are used for directed data mining, particularly classification. One of the chief advantages to decision trees is that the model is quite explainable since it takes the form of explicit rules (Berry, 1997:122).

A description of decision trees given by Pilot Software (1997b) is the following:

"Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include regression trees (CART) and Chi square Automatic Interaction Detection (CHAID)."

A description of Rule induction given by Pilot Software (1997b) is the following:

"The extraction of useful if-then rules from data based on statistical significance."

• Artificial Neural Networks

Neural networks are probably the most common data mining technique. They are simple models of neural interconnections in brains, adapted for use on digital computers. One of the chief advantages of neural networks is their wide applicability (Berry, 1997:122).
Another description of artificial neural networks is a non-linear predictive model that learns through training and resembles biological neural networks in structure (Pilot Software, 1997b).

- **Genetic Algorithms**

Genetic algorithms (GA) applies the mechanics of genetics and natural selection to a search used for finding the optimal sets of parameters that describe a predictive function. As such, it is used for directed data mining. Genetic algorithms are similar to statistics in that the form of the model needs to be known in advance. Genetic algorithms uses the selection, crossover, and mutation operators to evolve successive generations of solutions. As the generations evolve, only the most predictive survive until the functions converge on an optimal solution (Berry, 1997:123).

- **On-line Analytic Processing (OLAP)**

Not everything that is useful for data analysis is necessarily data mining. OLAP is a way of presenting relational data to facilitate understanding the data and important patterns inside it. Like visualisation, it is not specifically a tool for data mining, but it is an important tool in the arsenal of weapons used for extracting and presenting information (Berry, 1997:123).

The most common data mining techniques used for fraud detection and law enforcement is link analysis and clustering as illustrated by the actual case study below.
Davidson (1993:28) uses as example an actual fraud case and states the following:

When a major building society suspected it was the victim of an extensive fraud during the summer of 1987 it called in the police and eight policemen then spent two years combing through the evidence. Finally, they identified a suspect and made an arrest. The suspect confessed to running a complex operation in which, under different company names, he had acted as introducer, layer or accountant for a number of fraudulent mortgage applications. A two-year investigation is not excessive for this kind of case. In 1989, the Serious Fraud Office decided to find out whether a computerised technique for analysing data could speed up the process. It called a group of consultants with expertise in data analysis, gave them all the information it had used in its initial inquiry, and asked them to identify the criminal or his mode of operation. The consultants quickly got to work, first converting the information to consistent data formats and then checking the data for errors, ambiguities and repetitions. This took three days. They then fed the data into a software package called Netmap. Within minutes, the software highlighted the links between the fraudulent mortgage applications and the fraudster with his operating companies.

One of the Netmap graphs of the above-mentioned case is displayed as Figure 3.1 - Serious Fraud Office. The diagramme shows a trend between the introducer, accountant and bank. By visualising this trend/pattern, the undetected fraud was identified and the matter could be investigated immediately.

As can be seen from the above-mentioned example, using data mining techniques will make detecting fraud in organisations effective and efficient.
Figure 3.1 - Serious Fraud Office (Synectics Software, 1997)

Serious Fraud Office: 5 links or more
3.4 THE DATA MINING PROCESS - VIRTUOUS CYCLE (BERRY FRAMEWORK)

The promise of data mining is finding the interesting patterns lurking in all the billions and trillions of bytes of data. But merely finding patterns is not enough. You must be able to act on them, ultimately turning the data into information, the information into action, and the action into value. This is the virtuous cycle of data mining in a nutshell (Berry, 1997:18).

The Oxford Dictionary (1990:1371) defines a virtuous circle as follows: "It is a beneficial recurring cycle of cause and effect."

The four stages of the virtuous cycle of data mining are:
- Identify the business problem;
- Use data mining techniques to transform the data into actionable information;
- Act on the information; and
- Measure the results (refer to Figure 3.2).

Figure 3.2 - The four stages of the virtuous cycle of data mining (Berry, 1997:23)
As outlined in Figure 3.2, the key to success is incorporating data mining into business processes. These stages are highly interdependent: The results from one stage are the inputs into the next phase, much like the steps in a multi-step manufacturing process. The whole approach is driven by results. Each stage depends on the results from the previous stage (Berry, 1997:23).

Stage 1 - Identify the business opportunity

The purpose is to identify the areas where data can provide value. These, in turn, are the input in the Data Mining Stage. The goal is to identify areas where patterns in data have the potential of providing value (Berry, 1997:23).

Berry (1997:25) says that the information gathered in this step is critical to exploiting the virtuous cycle to its fullest.

In our scenario the business opportunity will be to detect fraud in the organisation and then to take the necessary steps to deal with it.

Stage 2 - Data Mining

This stage turns the data into information. Berry (1997:25) states that it takes data and business opportunities to produce actionable results for the Take Action Stage and that we need to understand what results it must produce to make the virtuous cycle successful.

Berry (1997:25) also identifies numerous pitfalls that can interfere with the ability to use the results of data mining. The more important pitfalls are:
• Bad data formats, such as using a five-digit zip code when nine-digit zip codes are needed by a direct mail application;
• Confusing data fields, such as a delivery date that means 'planned delivery date' in one system and 'actual delivery date' in another system;
• Lack of functionality, such as a call-centre application that does not allow annotations on a per-customer basis;
• Legal ramifications, such as having to provide a legal reason when rejecting a loan ("my neural network told me so" is not acceptable);
• Organisational factors, since some operational groups are reluctant to change their operations, particularly if they are not provided with incentives to change; and
• Timeliness, since results that come a month late may no longer be actionable.

Management and the auditor should take note of these pitfalls to manage the data mining process and to get the best results from their data.

Stage 3 - Take Action

After the data has been transformed into information, the organisation can take the necessary action which is to its benefit.

This is also where the results from data mining are acted upon and results are fed into the Measurement Stage (Berry, 1997:27).

Berry (1997:27) states that by requiring all business processes in an organisation to provide feedback before implementing any data mining solutions increases the risk of creating unnecessary delays. To avoid delays, plan for measuring the results a part of planning the initiative.
Stage 4 - Measure the results

Measurement provides the feedback for continuously improving results. Measurement refers specifically to measures of business value that go beyond response rates and costs, beyond averages and standard deviations. These measurements make the virtuous cycle of data mining virtuous (Berry, 1997:28).

Berry (1997:28) also states that it is a good idea to think of every data mining effort as a small business case. By comparing our expectations to the actual results, we can often recognise promising opportunities to exploit during the next round of the virtuous cycle.

The virtuous cycle of data mining is important as it gives an overview of the whole data mining (Knowledge Discovery in Databases) process, the different stages and the links between those stages.

3.5 THE KNOWLEDGE DISCOVERY PROCESS (ADRIAANS FRAMEWORK)

Adriaans (1996:37) states that the knowledge discovery process consists of six stages, namely data selection, cleaning, enrichment, coding, data mining and reporting and that the fifth stage, data mining, is the phase of real discovery.

The KDD (Knowledge Discovery in Databases) process is set out in Figure 3.3 on p.54.
Figure 3.3 is explained under the following headings:

- Information requirement;
- Data selection;
- Cleaning;
- Enrichment;
- Coding;
- Data mining;
- Reporting; and
- Action.
3.5.1 Information requirement

The starting point for any data mining activity is the formulation of a specific information requirement related to a specific action. That is, what do we want to know and what do we want to do with this knowledge? (Adriaans, 1996:81).

Adriaans (1996:82) states that apart from analysing marketing databases, there are several other areas where data mining can be very useful, such as analysis of large databases containing messages from helpdesks, analysis of management information bases produced by system management tools and analysis of data on financial transactions in order to detect fraud.

This short dissertation focuses on the fraud detection as information requirement.

3.5.2 Data selection

Adriaans (1996:83) states that once you have formulated your information requirements, the next logical step is to collect and select the data you need. In most cases, this data will be stored in operational databases used by the information systems in your organisation. However, gathering this information in a centralised database is not always an easy task since it may involve low-level conversion of data, such as from flat file to relational tables or from hierarchical systems to relational systems.

Each step in the KDD process is important, the data selection is an important step because if you select the wrong data set, the goal of fraud detection will not be achieved.
3.5.3 Cleaning

Adriaans (1996:84) states that once you have collected the data the next stage is cleaning. You are probably not aware of the amount of pollution that exists in your data. When databases are very large it is always advisable to select some random samples and analyse them to get a rough idea what one can expect. Almost all databases in large organisations are polluted and when we start to look at data from a data mining perspective, ideas concerning consistency of data change.

The cleaning stage is also important as the principle of 'garbage in, garbage out' still applies and the more polluted the data, the less probable it will be to identify the trends needed to detect fraud.

3.5.4 Enrichment

Once the data is cleaned, we may want to enrich it. In most countries access to many additional databases is available on a commercial basis and these can provide information on a variety of subjects, including demographic data such as the average prices of houses and cars, types of insurance that people have, and so on (Adriaans, 1996:85).

3.5.5 Coding

By means of selection and protection in SQL, we can manipulate data to obtain a clean target table. Sometimes we can remove pollution in the data simply by filtering out the polluted records. Suppose that some information concerning car or house ownership is lacking for some individuals in the database, if the lack of information is distributed randomly over the database, removing those records will not affect the type of clusters we will find (Adriaans, 1996:86).
3.5.6 Data mining

Data mining is not so much a single technique as the idea that there is more knowledge hidden in the data than shows itself on the surface. Any technique that helps extract more out of your data is useful, so data mining techniques form quite a heterogeneous group. Although various different techniques are used for different purposes, those that are of interest in the present context are:

- Query tools;
- Statistical techniques;
- Visualisation;
- Online Analytical Processing (OLAP);
- Case-based learning (k-nearest neighbour) also called MBR;
- Decision trees;
- Association rules;
- Neural networks; and

The more important data mining techniques were discussed in more detail in section 3.3.2 of this chapter.

The data mining stage is the stage where the data is transformed into information and the information into possible knowledge. The next step is to analyse the results of the trends/patterns and to extract the relevant knowledge to act upon.

3.5.7 Reporting

The reporting stage combines two different functions:

- **Analysis** of the results of the pattern recognition algorithms; and
- **Application** of the results of the pattern recognition algorithms.
We not only want to inspect what has been learned, but also apply the classifications and segmentation information that has been gathered. In many cases, reporting can be done using traditional query tools for databases. At present however, various new data visualisation techniques are emerging, ranging from simple scatter diagrams, showing different clusters into a two dimensional way, to complex interactive environments that enable us to fly over landscapes containing information about data sets (Adriaans, 1996:89).

3.5.8 Action

It is very important to take the necessary action after the discovery of knowledge. If this is not done, the whole process was fruitless.

Data mining should not be focussed on in isolation. Data mining is part of the knowledge discovery process and should be seen as such. Each step of the knowledge discovery process uses the information gathered in the preceding steps to achieve the set goals.

3.6 THE KNOWLEDGE DISCOVERY PROCESS (BRACHMAN & ANAND FRAMEWORK)

Knowledge discovery in databases is defined as follows by Fayyad (1996:6): "Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."

Fayyad (1996:7) states that the KDD process is a multi-step process which involves data preparation, search for patterns, knowledge evaluation, and refinement involving iteration after modification. The
process is assumed to be non-trivial - that is, to have some degree of search autonomy.

The KDD process is interactive and iterative, involving numerous steps with many decisions being made by the user. Brachman & Anand give a practical view of the KDD process (Fayyad, 1996:9-10).

Figure 3.4 outlines the KKD process of Brachman and Anand and the steps of turning raw data into knowledge.

Figure 3.4 - KDD process - Brachman & Anand (Fayyad, 1996:10)
The steps in the knowledge discovery process as outlined by Brachman and Anand in Fayyad (1996:10-11) as set out in Figure 3.4 (p.59) are:

- **Step 1**
  
  Developing an understanding of the application domain, the relevant prior knowledge and the *goals* of the end user.

- **Step 2**
  
  Creating a target data set *selecting* a data set or focusing on a subset of variables or data samples on which discovery is to be performed.

- **Step 3**
  
  *Data cleaning and pre-processing*. Basic operations such as the removal of noise or outliers if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, accounting for time sequence information and known changes.

- **Step 4**
  
  *Data reduction and Projection*. Finding useful features to represent the data depending on the goal of the task. Using dimensionality reduction or *transformation* methods to reduce the effective number of variables under consideration or to find invariant representations for the data.
• Step 5

Choosing the data mining task. Deciding whether the goal of the KDD process is classification, regression, clustering, etc.

• Step 6

Choosing the data mining algorithm(s). Selecting method(s) to be used for searching for patterns in the data.

• Step 7

Data mining. Searching for patterns of interest in a particular representational form or a set of such representations e.g. classification rules or trees, regression, clustering, etc.

• Step 8

Interpreting mined patterns, possible return to any of steps 1 - 7 for further iteration.

• Step 9

Consolidating discovered knowledge. Incorporating this knowledge into the performance systems or simply documenting it and reporting it to interested parties. This also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge (Fayyad, 1996:11).
The KDD process can involve significant iteration and may contain loops between any two steps. The basic flow of steps are illustrated in Figure 3.4 (p.59).

Most previous work on KDD has focused on step 7 (data mining). The other steps are however of considerable importance for the successful application of KDD in practice.

3.7 CONCLUSION

The evolution and techniques of data mining was discussed in this chapter. It was recognised that data mining is a part of the knowledge discovery in databases process and cannot be used in isolation. The knowledge discovery process was discussed in detail in this chapter and the main steps in the knowledge discovery process are the following:

- Step 1 - Information Requirement;
- Step 2 - Data Selection;
- Step 3 - Data Cleaning;
- Step 4 - Data Enrichment;
- Step 5 - Data Coding / Reduction;
- Step 6 - Data Mining;
- Step 7 - Interpretation / Reporting;
- Step 8 - Action; and
- Step 9 - Measurement and Evaluation.

The fraud detection process as discussed in chapter 2 resorts under Step 1 of the knowledge discovery process and takes place before the real knowledge discovery.

In Chapter 4 the knowledge discovery processes discussed in this chapter are combined to best fit the integrated fraud detection model.
Chapter 4: A MODEL FOR FRAUD DETECTION USING DATA MINING

Table of Contents

4.1 INTRODUCTION ................................................................. 64
4.2 OBJECTIVE OF THIS CHAPTER ........................................ 64
4.3 MODEL FOR FRAUD DETECTION USING DATA MINING....... 64
4.4 CONCLUSION ................................................................. 74

List of Figure

Figure 4.1 - A MODEL FOR FRAUD DETECTION USING
DATA MINING (KNOWLEDGE DISCOVERY) ....................... 65
4.1 INTRODUCTION

In this chapter the information gathered in chapter 2 and 3 is combined and an integrated model for fraud detection using data mining is developed which will assist management and auditors of an organisation to detect fraud in organisations. Management and auditors have a responsibility to detect fraud as discussed in chapter 2 of this short dissertation.

4.2 OBJECTIVE OF THIS CHAPTER

The objective of this chapter is to develop an integrated fraud detection model that uses data mining technology (KDD) to assist management and the auditor to detect fraud.

4.3 MODEL FOR FRAUD DETECTION USING DATA MINING

In chapter 2, section 2.8, the fraud detection process was described. In chapter 3 different knowledge discovery processes were described.

An integrated model for fraud detection using data mining was developed, as illustrated in Figure 4.1 (p.65) through combining the fraud detection process and the knowledge discovery processes.

The integrated fraud detection model (Figure 4.1) has three components. The components are the virtuous cycle, fraud detection process and products, which are discussed below.
Figure 4.1 - Integrated model for fraud detection using data mining (Knowledge Discovery)

Phase 1
- Identify business problems and areas where analysing data can provide value

Phase 2
- Transform data into actionable information using data mining techniques

Phase 3
- Act on the information

Phase 4
- Measure the results and efforts

Virtuous cycle

Feedback

Fraud detection process
- Identify the fraud symptoms indicators and red flags
- Identify the high fraud risk areas
- Identify common fraud areas in the organisation

Step 1(a)
Step 2
Step 3
Step 4
Step 5
Step 6
Step 7
Step 8
Step 9
Step 10

Operational data
Target data
Clean target data
Clean and enriched target data
Transformed target data
Fraud trends and patterns in data
Report on possible fraud trends and patterns
Results positive or negative
Improvement of process

Knowledge
4.3.1 Virtuous cycle

The promise of data mining is finding the interesting patterns lurking in all the billions and trillions of bytes of data. But merely finding patterns is not enough. You must be able to act on them, ultimately turning the data into information, the information into action, and the action into value. This is the *virtuous cycle* of data mining in a nutshell (Berry, 1997:18).

The four phases of the *virtuous cycle* of data mining as outlined in Figure 4.1 are:
- Identify the business problem;
- Use data mining techniques to transform the data into actionable information;
- Act on the information; and
- Measure the results and efforts.

For more detailed discussions regarding the virtuous cycle section 3.4 should be revisited.

The four phases in the virtuous cycle are relevant and applicable to the integrated fraud detection model. In each phase there is specific steps that a manager or auditor should follow to detect fraud in an organisation these steps are discussed in the next section.

4.3.2 Fraud detection process and products

Each phase with the relevant steps and the products of the integrated fraud detection model (*Figure 4.1*) is described below.
Phase 1 - Identify business problems

In this phase, management or the auditor should identify business problems and areas where analysing of the data can provide value. In our scenario the area where analysing data can add value is the area of fraud detection.

Fraud is a serious business problem in South Africa and the world and it is worthwhile for management and the auditor to make dealing with fraud a priority.

The steps of the first phase are discussed below:

> Step 1(a) - Identify the fraud symptoms, indicators and red flags

Management and auditors should know the fraud symptoms, indicators and red flags in an organisation. After each audit auditors should determine if any of the fraud symptoms, indicators or red flags were identified. This can be done using a checklist of the fraud symptoms, indicators or red flags. For a more detailed discussion on fraud symptoms, indicators and red flags see section 2.6.

> Step 1(b) - Identify common fraud areas in the organisation

This step could be done simultaneously with step nr. 1(a). The manager and auditor need to know the common fraud areas where fraud could occur in the organisation. Management and auditors should have a list of these areas and regularly update the list to be aware of the areas where fraud could occur in an organisation.
For a more detailed discussion on common fraud areas see section 2.7.

After following step 1(a) and 1(b) management and the auditors will be in a position to identify the high fraud risk areas.

> Step 2 - Identify the high fraud risk areas.

After step 1(a) and 1(b), management and the auditors should draw up a Business Fraud Risk Profile that will identify those areas in the organisation where the risk of fraud is high.

It is very important for management and the auditors to know which areas in the organisation are more susceptible to fraud in order to detect fraud in an organisation.

After the Business Fraud Risk Profile has been drawn up, managers or auditors need to identify any data available for analysing in the high fraud risk area. If there is data, they can go to the next phase. If there is no data available, consideration should be given to detect the fraud manually.

The product, having followed the steps in phase one, is a Business Fraud Risk Profile. The Business Fraud Risk Profile is a document that management or the auditor will prepare to identify the high, medium and low fraud risk areas. Management or the auditor should then concentrate their efforts on the high fraud risk areas to detect fraud.
Phase 2 - Transform data into actionable information using data mining techniques

In this phase raw data is identified and turned into information and the information turned into knowledge.

In this phase there are six steps to follow:
- Data selection;
- Data Cleaning;
- Data Enrichment;
- Data Coding/reduction;
- Data Mining; and
- Interpretation/Reporting.

The steps and products of the second phase are discussed in more detail below:

Step 3 - Data selection

After the high fraud risk areas in the organisation have been identified, the next step is to identify the relevant data for detecting the fraud in the organisation.

The data could be in different locations, e.g. a data warehouse, on a server, on a personal computer, etc. It is very important to identify the relevant data, because if the wrong data is identified, the probability to detect a fraud will decline. For a more detailed discussion on data selection see section 3.5.2.

The product that will be identifiable after this step has been performed is that management or the auditor will have selected
the relevant operational data (target data) to start the knowledge discovery process. This target data would be the data that is relevant to the high fraud risk area identified in the previous step.

The next step is to clean the relevant target data.

➢ Step 4 - Data Cleaning

After the selection, the target data needs to be cleaned. Obvious errors in the data need to be corrected e.g. a date that does not exist 1999-99-99.

The cleaning stage is also important as the principle of 'garbage in garbage out' still applies and the more polluted the data, the less probable it will be to identify the trends needed to detect fraud. For a more detailed discussion on data cleaning see section 3.5.3.

The product that will be identifiable after this step has been performed is that management or the auditor will have relevant clean target data.

The next step is to enrich the relevant and clean target data.

➢ Step 5 - Data Enrichment

After cleaning the data, management and the auditor need to identify any similar or other relevant data that should be added to the current data to improve the probability to detect fraud. For a more detailed discussion on data enrichment see section 3.5.4.
The product that will be identifiable after this step has been performed is that management or the auditor will have relevant clean and enriched target data.

➢ Step 6 - Data Coding / Reduction

Before the data mining can start, the manager or auditor need to identify the fields of data that they are going to use in the data mining study. They can use tools like SQL to get a clean set of target data that will be ready to mine for trends and patterns to identify fraud. For a more detailed discussion on data coding see section 3.5.5.

The product that will be identifiable after this step has been performed is that management or the auditor will have relevant, transformed, enriched and clean target data.

The next step is to mine the relevant, transformed, enriched and clean target data.

➢ Step 7 - Data Mining

In this step, management and auditors can use data mining techniques (software tools) to mine for trends and patterns that will point them to the possible fraudulent transactions in the data. For a more detailed discussion on data mining see section 3.5.6.

The product that will be identifiable after this step has been performed is the possible fraud trends and patterns in the data that management or the auditor have identified.
The next step is to interpret and report the identified trends and patterns.

➤ **Step 8 - Interpretation / Reporting**

After the trends and the patterns have been identified, management and the auditor can interpret the patterns in the data that could possibly point towards fraud.

These trends and patterns should be reported to the relevant parties for further follow-up/take-action. For a more detailed discussion on reporting see section 3.5.7.

The **product** that will be identifiable after this step has been performed is the report issued to the relevant parties on the possible fraud trends and patterns in the data that management or the auditor have identified.

The next step is to act on the trends and patterns in the data.

- **Phase 3 - Act on the information**

At this stage the data has been transformed into information and the information should be acted upon. The only step in the third phase is:

➤ **Step 9 - Action**

The reported possible fraud needs to be followed up and the results could be either positive (fraud detected) or negative (no fraud detected).
If a fraud is detected, management can then take the corrective action that is needed, e.g. investigation of fraud, implementing of necessary controls and prevention programmes throughout the organisation.

The product that will be identifiable after this step has been performed is that fraud was either discovered or no fraud was discovered. Results could either be positive or negative. The corrective action taken by management will also be a 'product'.

- Phase 4 - Measure the results and effort

The only step in the fourth phase is:

➢ Step 10 - Measurement and evaluation

We have to measure and evaluate our positive or negative results and then use the measurement as input in the first phase of this fraud detection model.

The product that will be identifiable after this step has been performed is that the process of fraud detection can be improved with the information gathered during the fraud detection process.

4.3.3 General

All the above-mentioned steps are interdependent as shown in Figure 4.1. Regular feedback should be given between stages to improve the fraud detection process.
4.4 CONCLUSION

An integrated model for fraud detection using data mining (knowledge discovery) was developed which will assist management and the auditor to understand fraud detection and to detect fraud.

Management and the auditors need to understand that detecting fraud is not an easy task. If the above mentioned model is followed it would however help them to detect fraud in organisations.
Chapter 5: SUMMARY OF RESULTS AND CONCLUSION

Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1  SUMMARY OF RESULTS</td>
<td>76</td>
</tr>
<tr>
<td>5.2  OVERALL CONCLUSION</td>
<td>77</td>
</tr>
<tr>
<td>5.3  FURTHER RESEARCH</td>
<td>77</td>
</tr>
</tbody>
</table>
5.1 SUMMARY OF RESULTS

5.1.1 Fraud detection

Fraud is a serious business problem in South Africa and the world. Management and auditors need a model to detect fraud in organisations as they have a responsibility to detect fraud and they all have a role to play even if their responsibilities differ. An effective and efficient way of detecting fraud is by using a technology called data mining which is part of the knowledge discovery in databases process.

A fraud detection process was developed and steps in the process were identified as the following:

- Step 1 - Identify the fraud symptoms, indicators and red flags in the organisation;
- Step 2 - Identify the common fraud areas in the organisation;
- Step 3 - Identify the high fraud risk areas; and
- Step 4 - Audit for fraud by using a technology called data mining (Knowledge Discovery Process - discussed in chapter 3).

5.1.2 Data mining

Data mining is a technology that is part of the Knowledge discovery process in databases.

The steps in the knowledge discovery in databases process is the following:

- Step 1 - Information requirement;
- Step 2 - Data Selection;
- Step 3 - Data Cleaning;
• Step 4 - Data Enrichment;
• Step 5 - Data Coding / Reduction;
• Step 6 - Data Mining;
• Step 7 - Interpretation / Reporting;
• Step 8 - Action; and
• Step 9 - Measurement and Evaluation.

5.1.3 Model for fraud detection using data mining

The need for a model to assist management and auditors was identified. A model was developed as discussed in detail in chapter 4. By using the technology called data mining fraud can be detected more effectively and efficiently by management and auditors.

5.2 OVERALL CONCLUSION

The objectives, as stated in chapter 1, have been reached:
An integrated model for fraud detection using data mining (included in the knowledge discovery process) has been developed which will assist management and auditors to detect fraud and will add value to organisations that will be able to detect fraud in a more effective and efficient way.

5.3 FURTHER RESEARCH

An area for further research would be to explore the other stages of the process of dealing with fraud as discussed in section 1.3.2, e.g. prevention, and link them to a relevant technology that can assist management and the auditor in their fight against fraud.
BIBLIOGRAPHY


FRAUD DETECTION USING DATA MINING

FAYYAD, U M; PIATETSKY-SHAPIRO, G; SMYTH, P & UTHURUSAMY, R


GRANT THORNTON, 1997: Auditing Standards Board proposes new fraud Detection standards. [Online]. Available URL address:
http://gt.com/gtonline/assuranc /pcady/pcasumb.html


PRICE, D 1991: Fraudbusting – How to identify and deal with Corporate Fraud and how to prevent it. Mercury.


SAICA, 1992: AU005, The auditor's responsibility to detect and report illegal acts, other irregularities and errors in the Member's Handbook (South African Institute of Chartered accountants, JHB)

SYNECTICS SOFTWARE, 1997: Serious fraud office : 5 links and more (Unpublished)
