DATA –DRIVEN FRAUD DETECTION

SUBMITTED BY

MANOJ SHARMA
AMIT
SANKET

Errors and Frauds

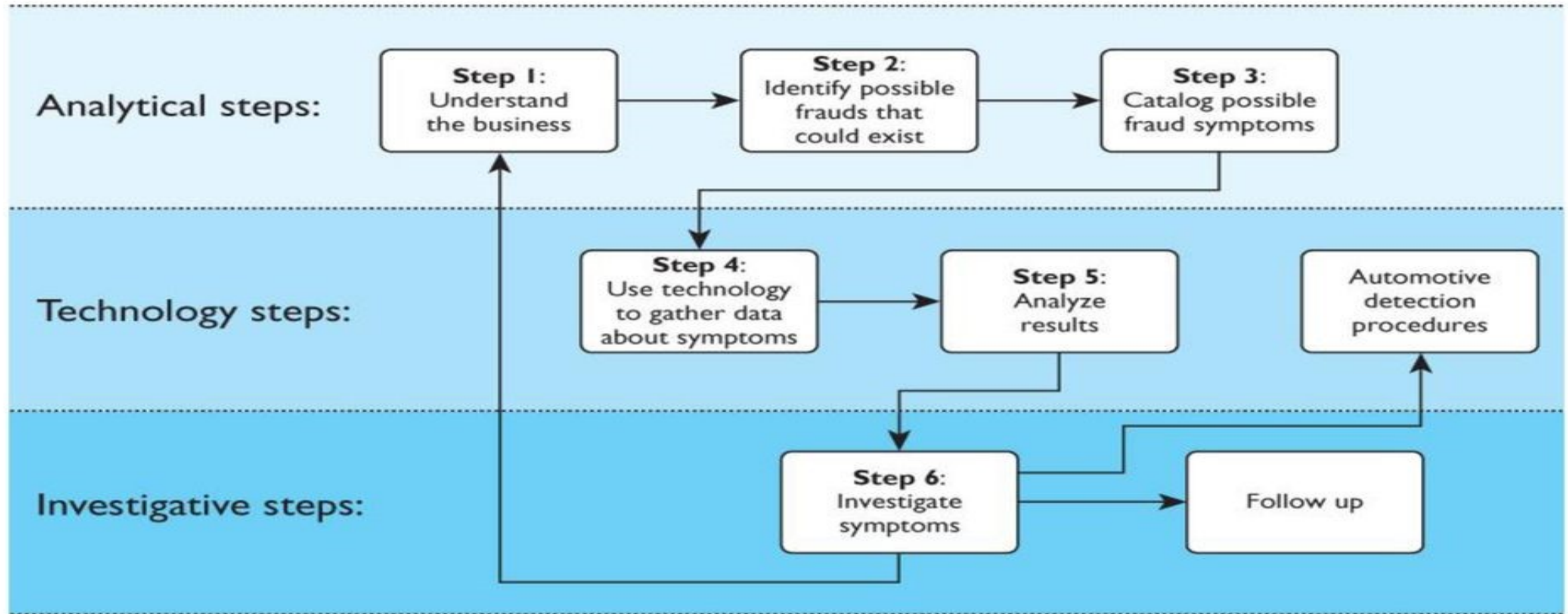| ERRORS	FRAUD
▶ Errors are not intentional ▶ They are simply problems in the system caused by failures in systems, procedures, or policies ▶ They do not represent fraud and normally do not result in legal action ▶ Errors are usually spread evenly throughout a data set	▶ Fraud is the intentional circumvention of controls by intelligent human beings ▶ Perpetrators cover their tracks by creating false documents or changing records in database systems ▶ Evidence of fraud may be found in very few transactions ▶ Fraudulent symptoms are found in single cases or limited areas of the data set

Audit Sampling and Fraud

▶ Statistical sampling has become a standard auditing procedure.
  ▶ Audit sampling is an effective analysis procedure for finding routine errors spread throughout a data set.
  ▶ In contrast, sampling is usually a poor analysis technique when looking for a needle in a haystack.
  ▶ If you sample at a 5 percent rate, you effectively take a 95 percent chance that you will miss the few fraudulent transactions.

▶ Often, fraud examiners strive to complete full-population analysis to ensure that the "needles" are found.

▶ Given the right tools and techniques, full-population analysis is often the preferred method in a fraud investigation.

The Data Analysis Process

▶ Fraud investigators must be prepared to learn new methodologies, software tools, and analysis techniques to successfully take advantage of data-oriented methods.

▶ Data-driven fraud detection is proactive in nature.
  ▶ The investigator no longer has to wait for a tip to be received.
  ▶ The investigator brainstorms the schemes and symptoms that might be found and then looks for them.

▶ Data-driven detection is essentially a hypothesis-testing approach:
  ▶ The investigator makes hypotheses and tests to see which are supported by the data.

Figure 6.1 The Proactive Method of Fraud Detection

Analytical steps:

Step 1: Understand the business → Step 2: Identify possible frauds that could exist → Step 3: Catalog possible fraud symptoms

Technology steps:

Step 4: Use technology to gather data about symptoms → Step 5: Analyze results

Automotive detection procedures

Investigative steps:

Step 6: Investigate symptoms → Follow up

The Data Analysis Process—Six Steps

Step 1 Understand the business

Step 2 Identify possible frauds that could exist

Step 3 Catalog possible fraud symptoms

Step 4 Use technology to gather data about symptoms

Step 5 Analyze results

Step 6 Investigate symptoms

Step 1: Understand the Business

▶ The same fraud detection procedures cannot be applied generically to all businesses or even to different units of the same organization.

▶ Several potential methods to gather information about a business are as follows:

  ▶ Tour the business, department, or plant
  ▶ Become familiar with competitor processes
  ▶ Interview key personnel (ask them where fraud might be found)
  ▶ Analyze financial statements and other accounting information
  ▶ Review process documentation
  ▶ Work with auditors and security personnel
  ▶ Observe employees performing their duties

Step 2: Identify Possible Frauds That Could Exist

▶ This risk assessment step requires an understanding of the nature of different frauds, how they occur, and what symptoms they exhibit.

▶ The fraud identification process begins by conceptually dividing the business unit into its individual functions or cycles.

▶ During this stage, the fraud detection team should brainstorm potential frauds by type and player.

Step 3: Catalog Possible Fraud Symptoms

▶ In Step 3, the fraud examiner should carefully consider what symptoms could be present in the potential frauds identified in Step 2.

Types of Fraud Symptoms

- ▶ Accounting errors
- ▶ Internal control weaknesses
- ▶ Analytical errors
- ▶ Extravagant lifestyles
- ▶ Unusual behaviors
- ▶ Tips and complaints

Step 4: Use Technology to Gather Data about Symptoms

- Searching and analysis
  - Data analysis applications
  - Custom structured query language (SQL) queries and scripts
- The deliverable of this step is a set of data that matches the symptoms identified in the previous step.

Step 5: Analyze Results

- Once errors are refined and determined by the examiners to be likely indications of fraud, they are analyzed using either traditional or technology-based methods:
  - Screening results using computer algorithms
  - Real-time analysis and detection of fraud
- One advantage of the data-driven approach is its potential reuse.

Step 6: Investigate Symptoms

- The final step of the data-driven approach is investigation into the most promising indicators.
- The primary advantage of the data-driven approach is the investigator takes charge of the fraud investigation process.
  - Instead of waiting for tips or other indicators to become egregious enough to show on their own, the data-driven approach can highlight frauds while they are small.
- The primary drawback to the data-driven approach is that it can be more expensive and time intensive than the traditional approach.

Text Import

- Several text formats exist for copying data from one application (i.e., a database) to another (i.e., an analysis application).
  - Delimited text
    - Comma separated values (CSV)
    - tab separated values (TSV)
  - Fixed-width format
  - Extensible markup language (XML)
    - Used in many new applications
  - EBCDIC
    - Used primarily on IBM mainframes

Data Analysis Techniques

▶ Once data are retrieved and stored in a data warehouse, analysis application, or text file, they need to be analyzed to identify transactions that match the indicators identified earlier in the process.

▶ Analysis techniques commonly used by fraud investigators:
  ▶ Data preparation
  ▶ Benford's Law
  ▶ Digital analysis
  ▶ Outlier investigation
  ▶ Stratification and summarization
  ▶ Time trend analysis
  ▶ Fuzzy matching
  ▶ Real-time analysis

Summarization

▶ Summarization is an extension of stratification.

▶ Summarization runs one or more calculations on the subtables to produce a single record representing each subtable.

▶ Basic summarization usually produces a single results table with one record per case value.

▶ Pivot tables (also called cross tables) are two-dimensional views with cases in one dimension and the calculations in the detail cells.

Real-Time Analysis

▶ Data-driven investigation is one of the most powerful methods of discovering fraud.

▶ It is usually performed during investigations or periodic audits, but it can be integrated directly into existing systems to perform real-time analysis on transactions.

▶ Although real-time analysis is similar to traditional accounting controls because it works at transaction time, it is a distinct technique because it specifically analyzes each transaction for fraud (rather than for accuracy or some other attribute).