2018 IIA INDONESIA NATIONAL CONFERENCE

Nurturing Agile Internal Auditors in Disruptive Times

FRAUD DETECTION WITH DATA ANALYTIC & MACHINE LEARNING

Fandhy H. Siregar, M. Kom., CIA, CRMA, CCSA, CISA, CISM, CRISC, CGEIT, CISSP, CEH, COBIT5, CEP-PM, QIA

Last presented in:

2017
ASIA PACIFIC
CACS™
AN ISACA EVENT
29 – 30 November | Dubai

The Institute of Internal Auditors Indonesia

2018 NATIONAL CONFERENCE
Indonesia Bali, 28-29 August

CHAIR

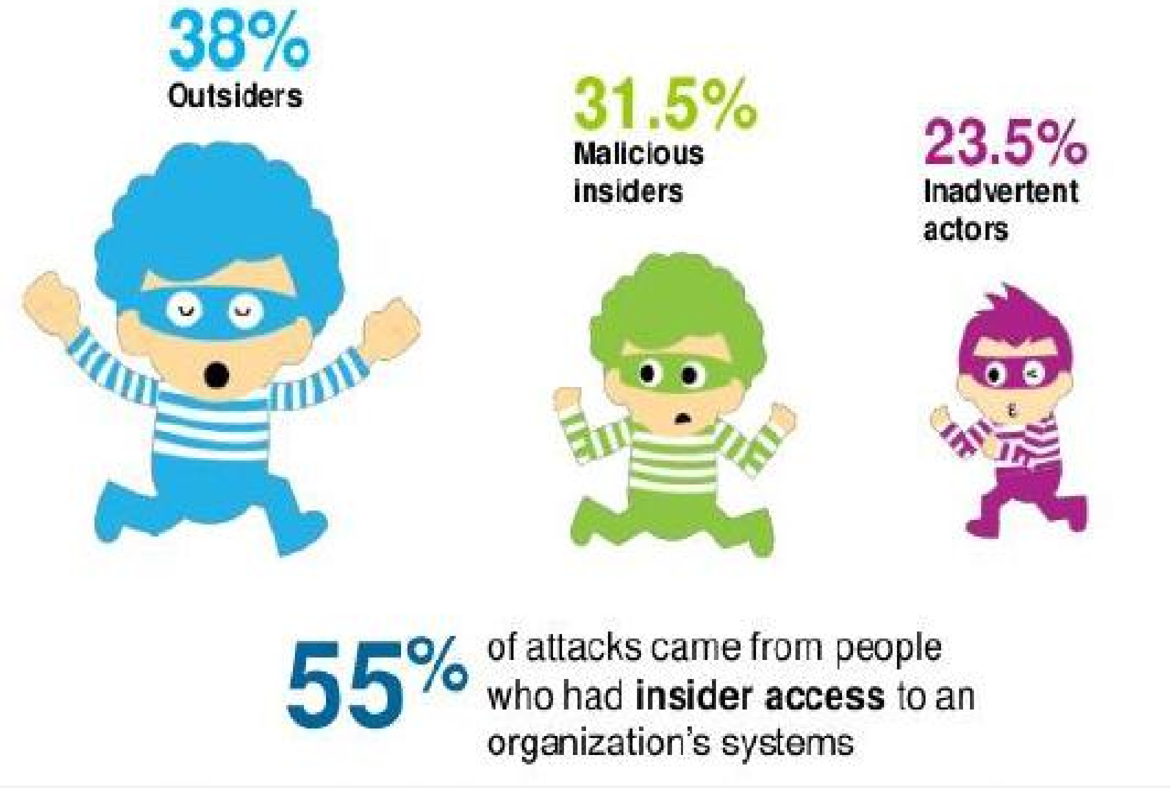
Rama Kurnia

Governor IIA Indonesia
Auditor Internal SKK Migas

Fandhy Haristha Siregar

Secretary IIA Indonesia
CAE Bank Resona Perdania

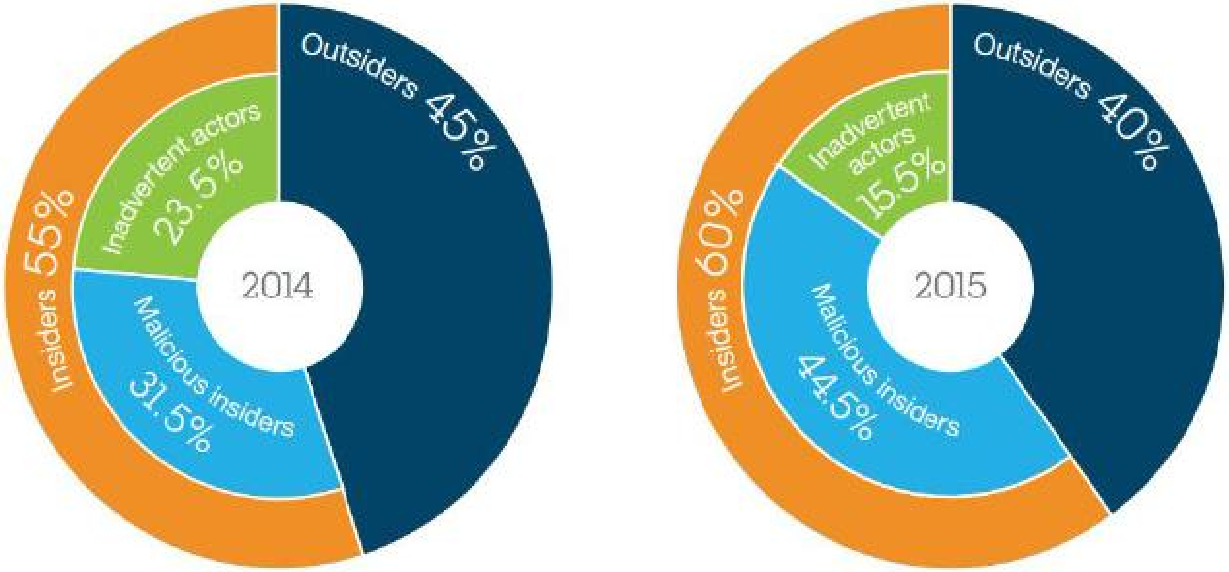
EXTERNAL VS INTERNAL THREATS



55% of attacks came from people who had **insider access** to an organization's systems

Source: IBM ® X-Force ® Research

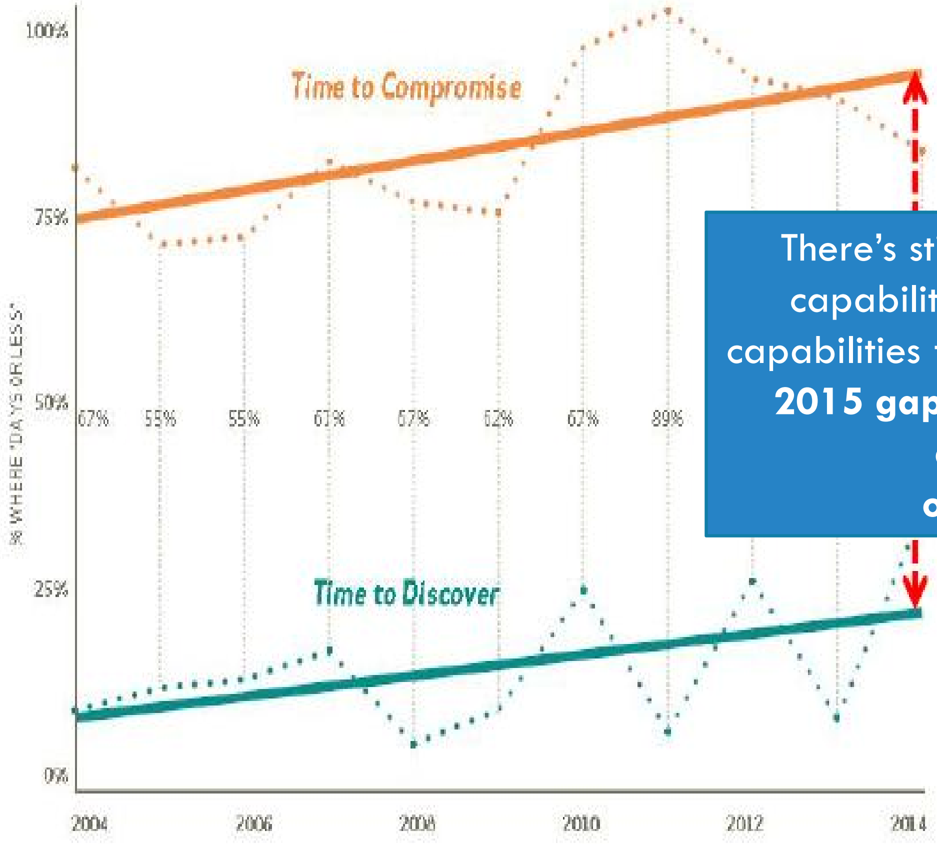
EXTERNAL VS INTERNAL THREATS



In 2015, **60 percent** of all attacks were carried out by **insiders**, either ones with malicious intent or those who served as **inadvertent actors**. In other words, they were instigated by people you'd be likely to trust. And they can result in **substantial financial and reputational losses**.

Source: IBM ® X-Force ® Research

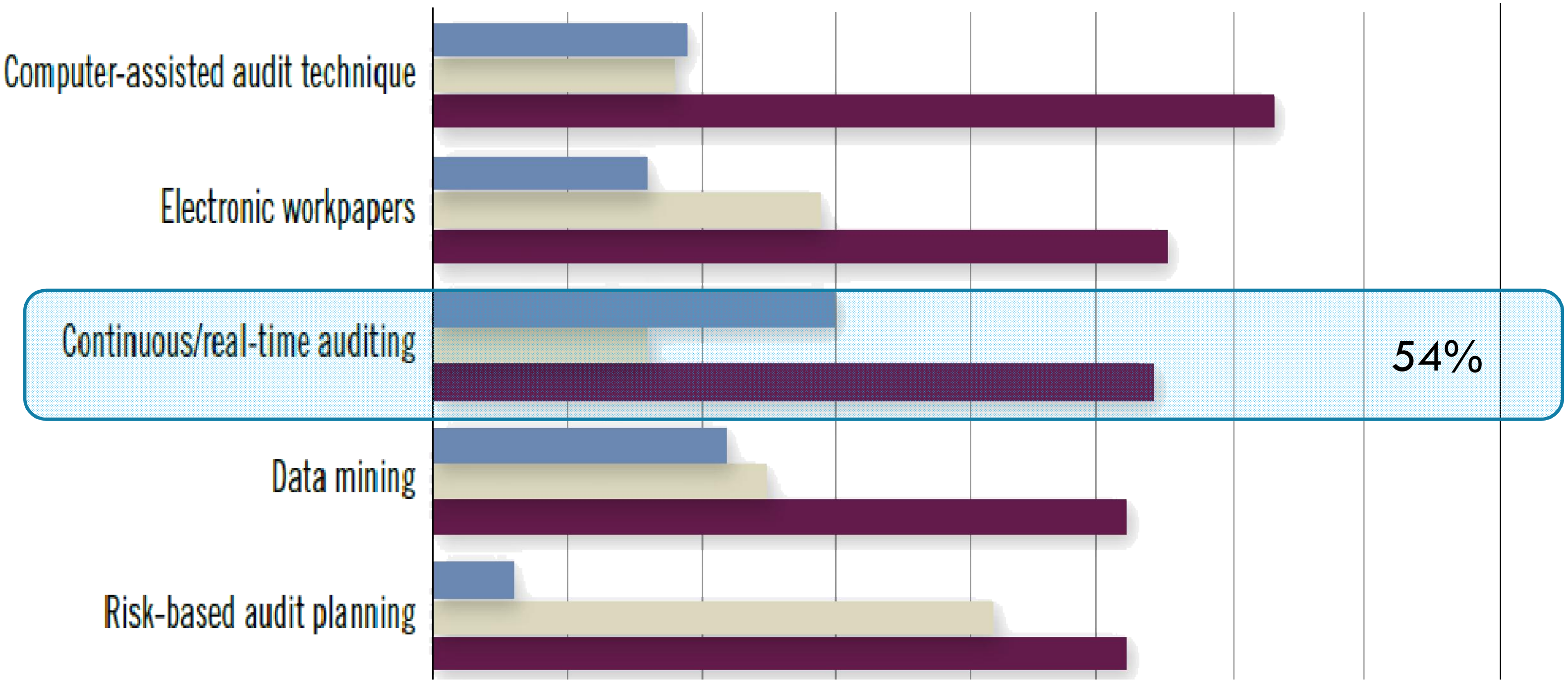
DETECTION GAP



There's still a gap between capabilities to detect with capabilities to deliver the attacks
2015 gap was jumping up again to
over 60%

Source: VerizoneData Breach Investigation Report (DBIR)

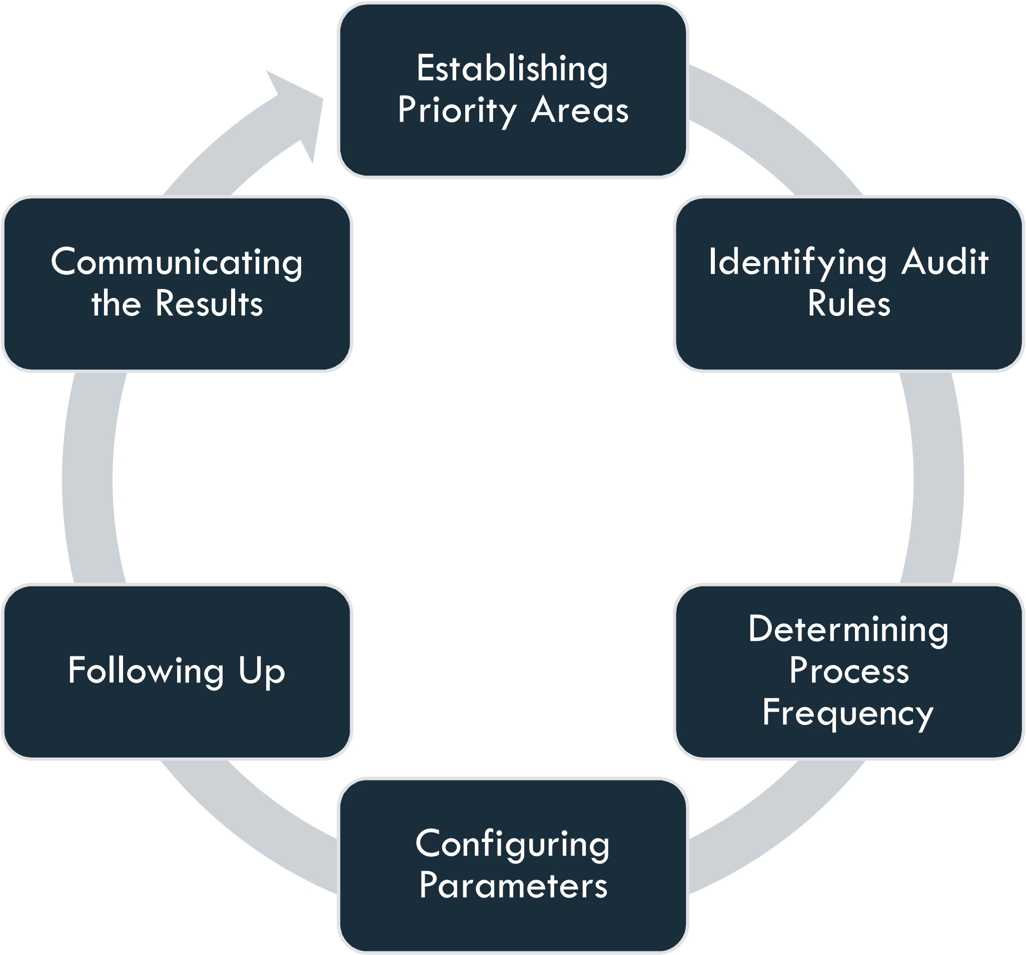
TOP 5 AUDIT TOOLS



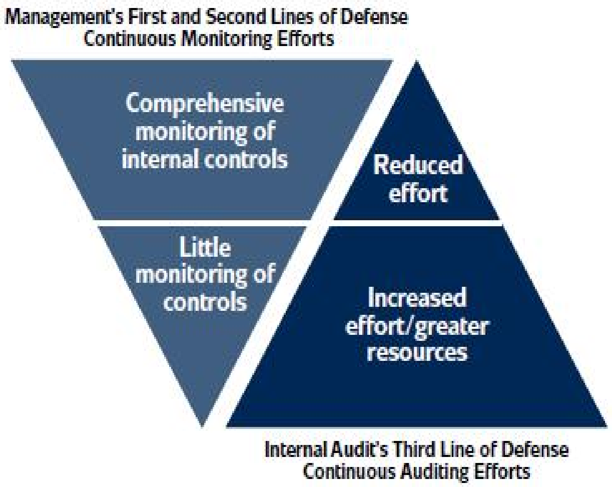
🟦 Not be used and less than now ⬜ Be used about the same as now 🟥 Be used and more than now

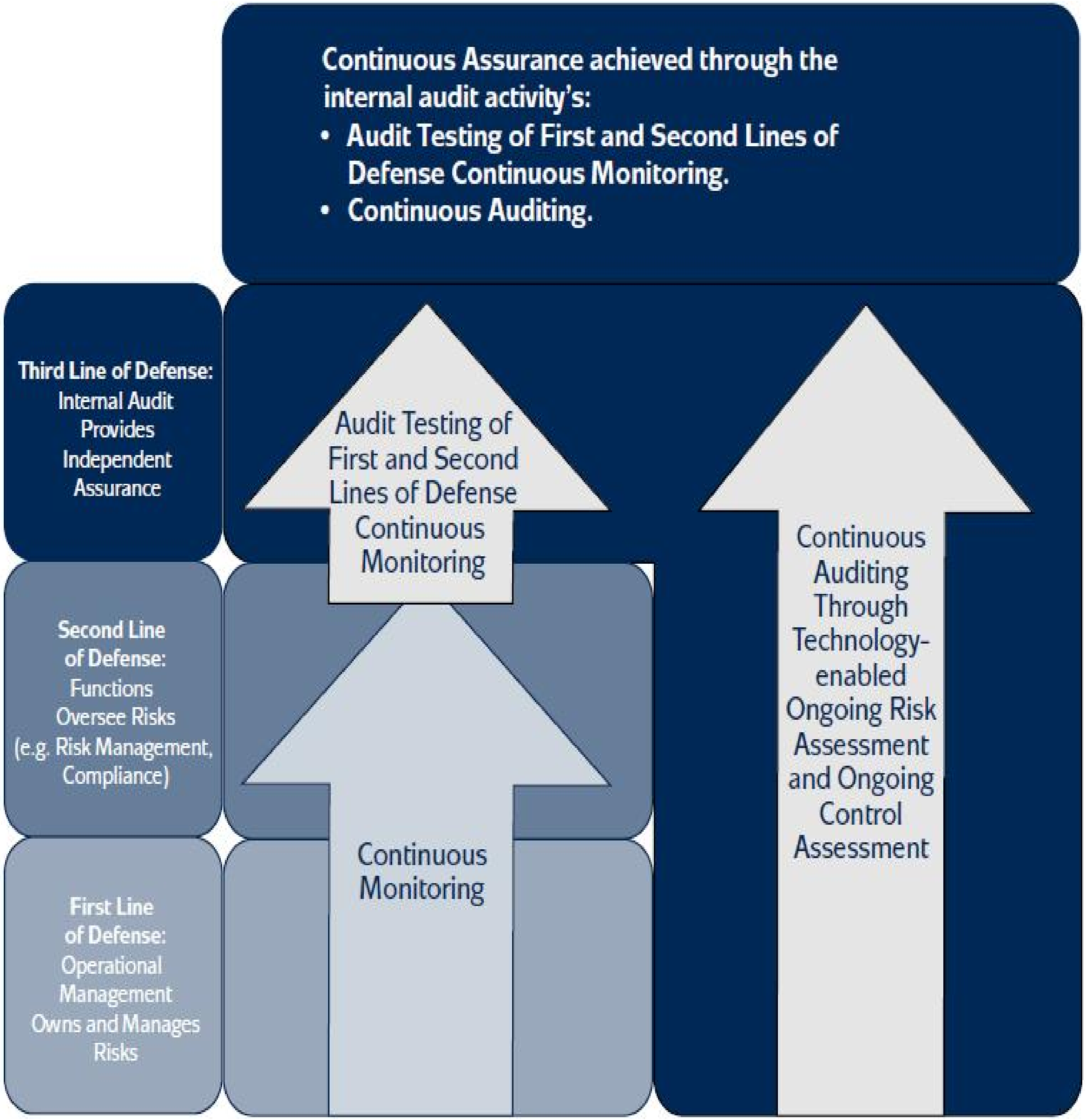
Source: Global Technology Audit Guide (GTAG), IIA

CONTINUOUS AUDIT IMPLEMENTATION STEPS


Establishing Priority Areas
Identifying Audit Rules
Determining Process Frequency
Configuring Parameters
Following Up
Communicating the Results

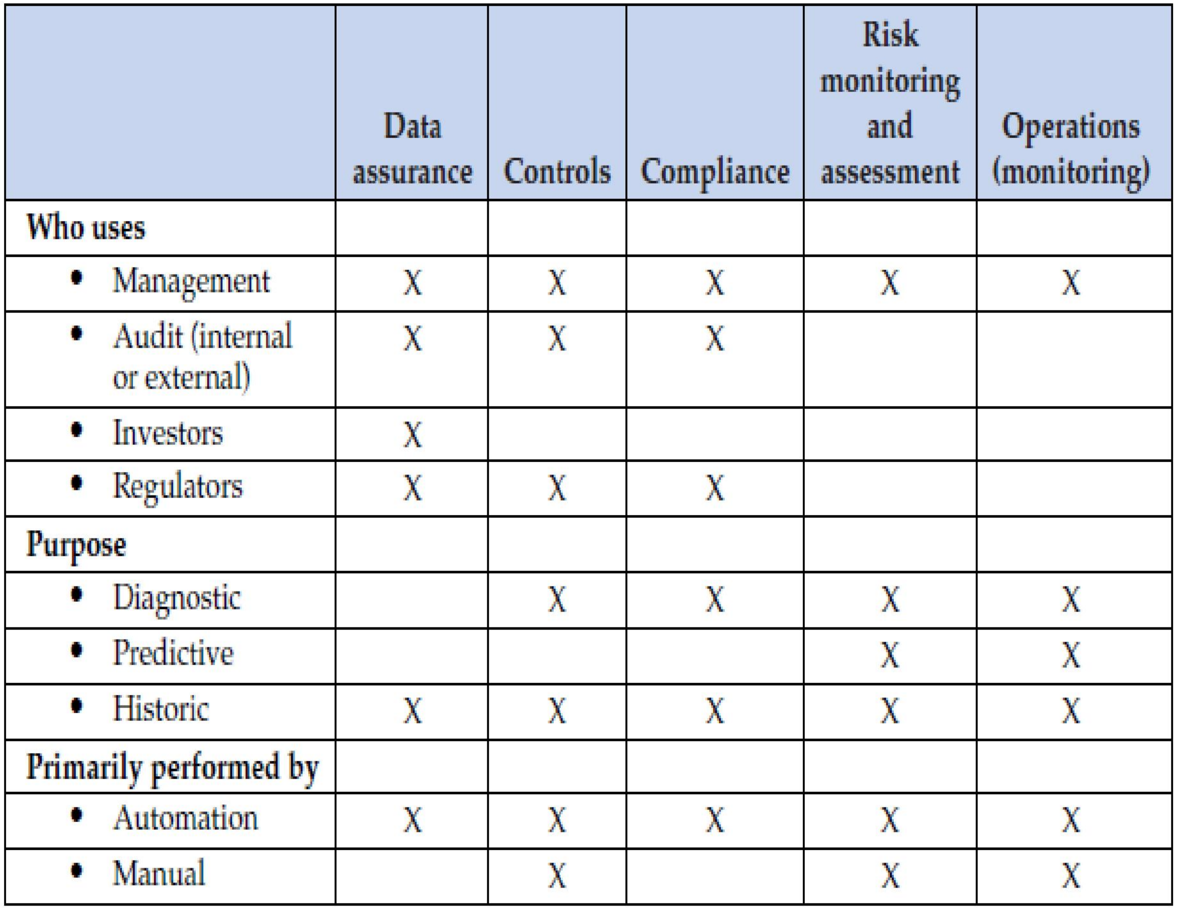
CONTINUOUS AUDITING VS CONTINUOUS MONITORING


Management's First and Second Lines of Defense Continuous Monitoring Efforts
Comprehensive monitoring of internal controls
Reduced effort
Little monitoring of controls
Increased effort/greater resources
Internal Audit's Third Line of Defense Continuous Auditing Efforts


Continuous Assurance achieved through the internal audit activity's:
• Audit Testing of First and Second Lines of Defense Continuous Monitoring.
• Continuous Auditing.
Third Line of Defense: Internal Audit Provides Independent Assurance
Second Line of Defense: Functions Oversee Risks (e.g. Risk Management, Compliance)
First Line of Defense: Operational Management Owns and Manages Risks
Audit Testing of First and Second Lines of Defense Continuous Monitoring
Continuous Monitoring
Continuous Auditing Through Technology-enabled Ongoing Risk Assessment and Ongoing Control Assessment

Source: Global Technology Audit Guide (GTAG), IIA

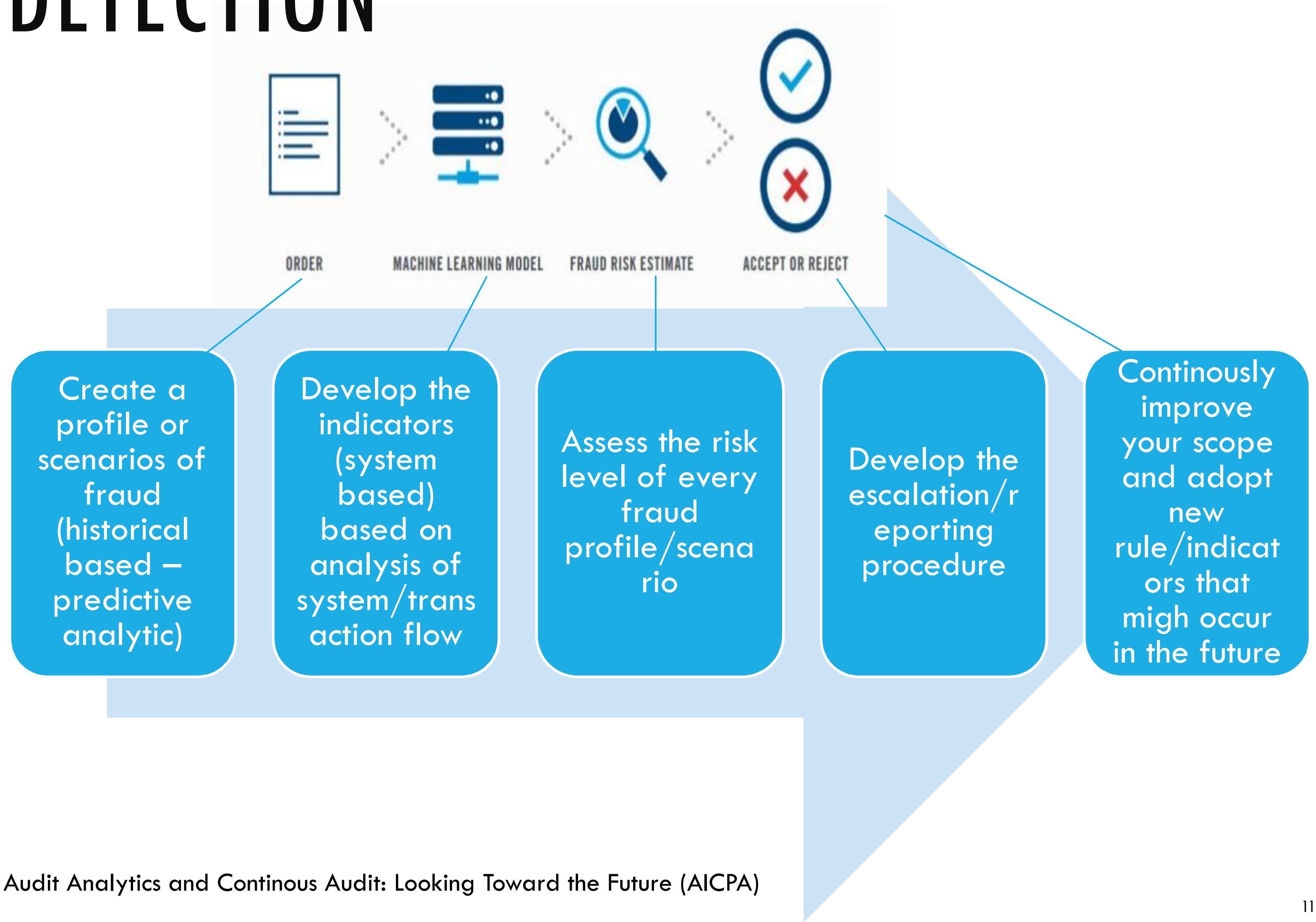
CONTINUOUS COMBINED ASSURANCE



	Data assurance	Controls	Compliance	Risk monitoring and assessment	Operations (monitoring)
Who uses					
• Management	X	X	X	X	X
• Audit (internal or external)	X	X	X		
• Investors	X				
• Regulators	X	X	X		
Purpose					
• Diagnostic		X	X	X	X
• Predictive				X	X
• Historic	X	X	X	X	X
Primarily performed by					
• Automation	X	X	X	X	X
• Manual		X		X	X

Source: Audit Analytics and Continous Audit: Looking Toward the Future (AICPA)

BUILDING YOUR OWN FRAUD SELF-DETECTION



Create a profile or scenarios of fraud (historical based – predictive analytic)

Develop the indicators (system based) based on analysis of system/transaction flow

Assess the risk level of every fraud profile/scenario

Develop the escalation/reporting procedure

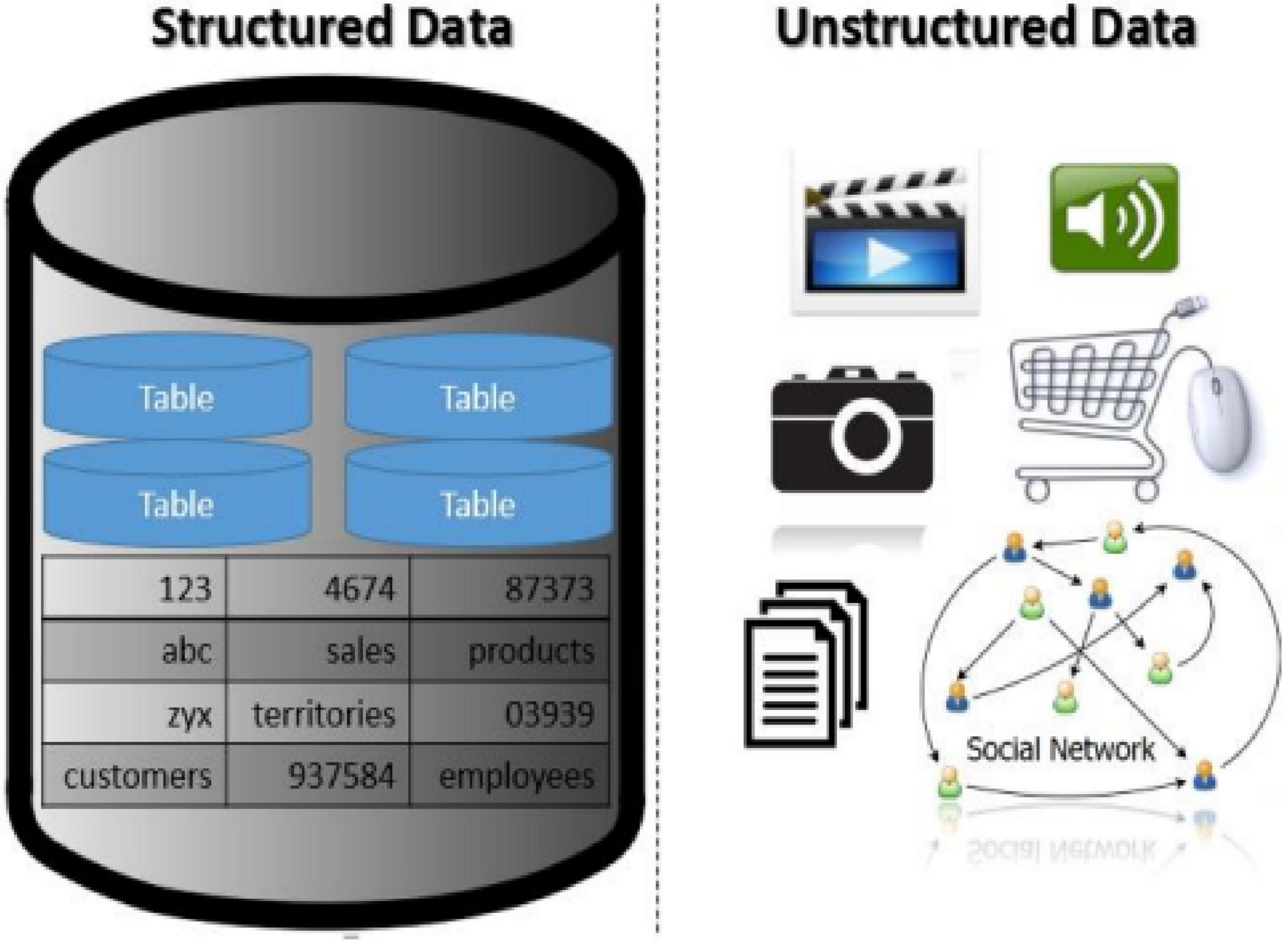
Continously improve your scope and adopt new rule/indicators that migh occur in the future

Source: Audit Analytics and Continous Audit: Looking Toward the Future (AICPA)

DATA ANALYTICS

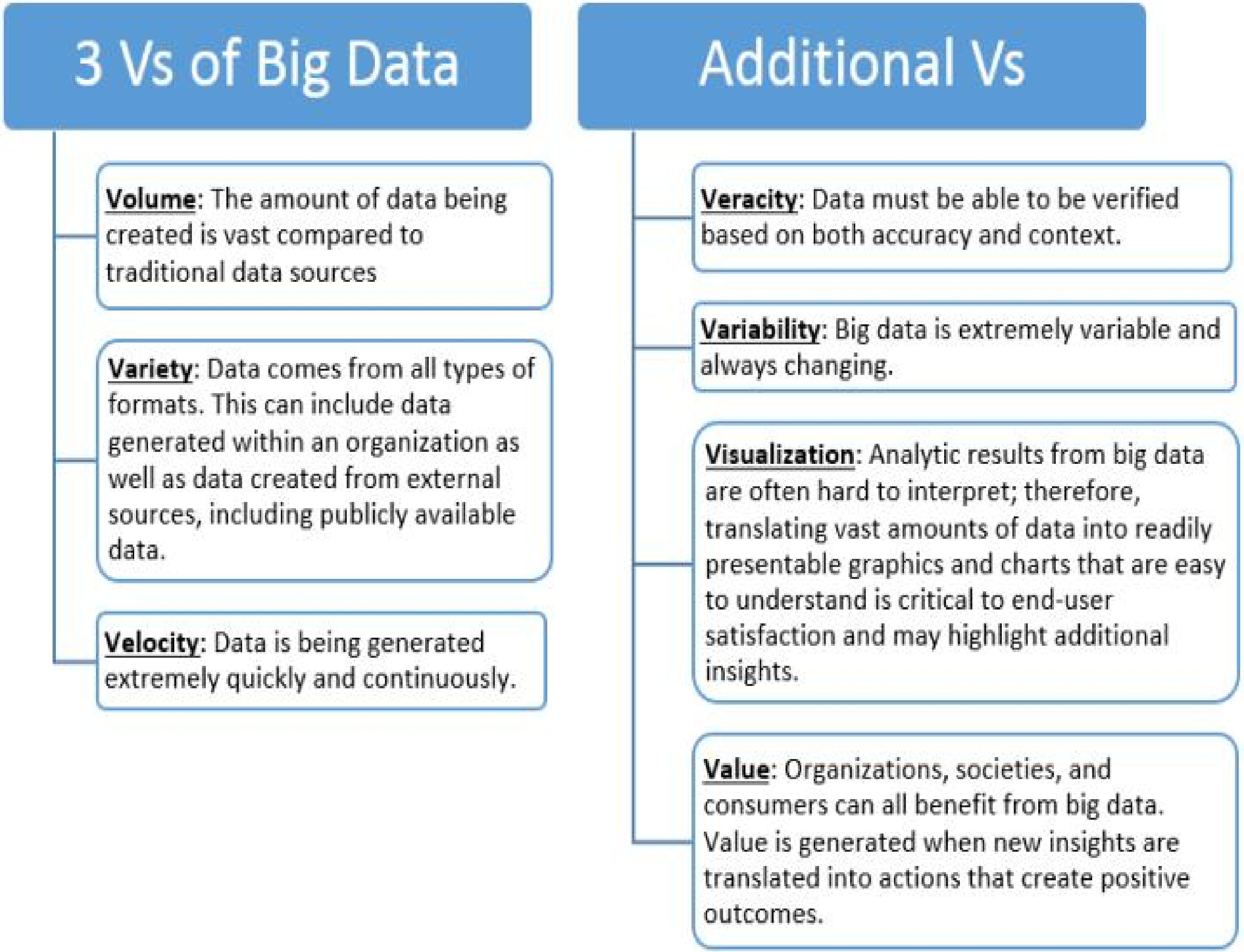
- Data analytics, as it applies to fraud examination, refers to the use of analytic software to identify trends, patterns, anomalies, and exceptions in data.
- Especially useful when fraud is hidden in large data volumes and manual checks are insufficient.
- Can be reactively and proactively
- Data analysis techniques alone are unlikely to detect fraud; human judgment is needed to decipher results.
- To detect fraud, data analysis techniques must be performed on the full data population, instead of using sample of data.

TYPE OF DATA THAT CAN BE ANALYZED



123	4674	87373
abc	sales	products
zyx	territories	03939
customers	937584	employees

BIG DATA AUDITING



3 Vs of Big Data

Volume: The amount of data being created is vast compared to traditional data sources

Variety: Data comes from all types of formats. This can include data generated within an organization as well as data created from external sources, including publicly available data.

Velocity: Data is being generated extremely quickly and continuously.

Additional Vs

Veracity: Data must be able to be verified based on both accuracy and context.

Variability: Big data is extremely variable and always changing.

Visualization: Analytic results from big data are often hard to interpret; therefore, translating vast amounts of data into readily presentable graphics and charts that are easy to understand is critical to end-user satisfaction and may highlight additional insights.

Value: Organizations, societies, and consumers can all benefit from big data. Value is generated when new insights are translated into actions that create positive outcomes.

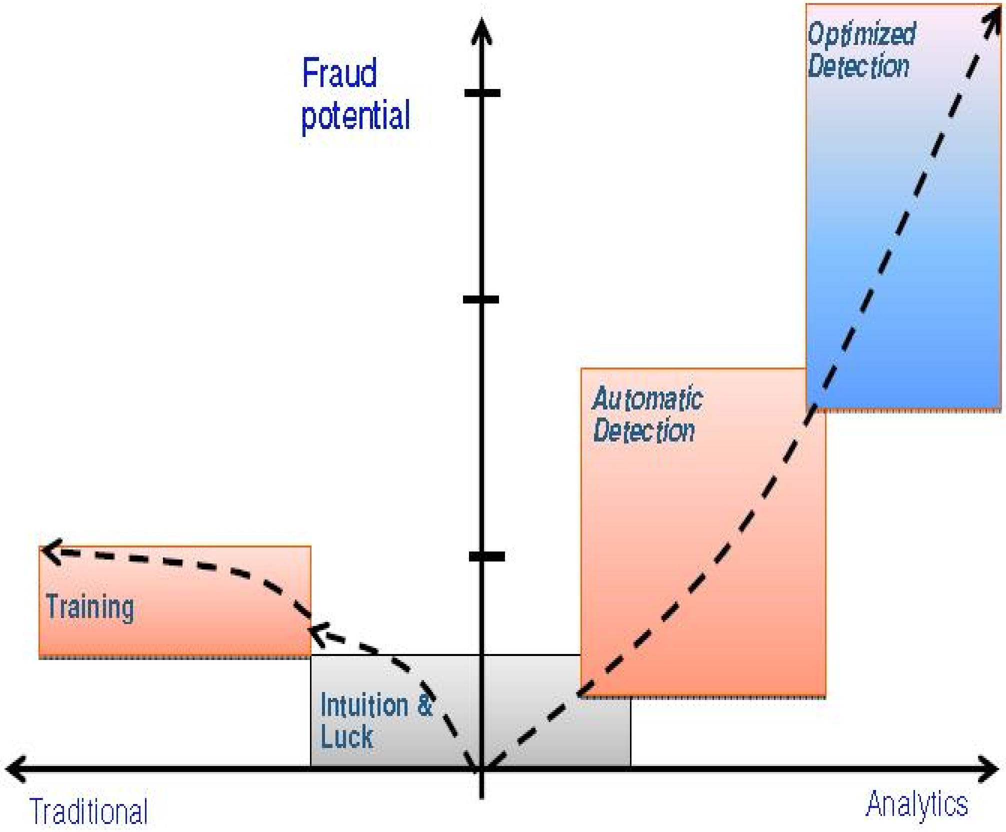
Source: Understanding and Auditing Big Data, Global Technology Audit Guide (GTAG), IIA

OUNCE OF PREVENTION = POUND OF CURE

- A big part of fraud prevention is communicating the program across the organization.
- If everyone knows there are systems in place that alert to potential fraud or breach of controls, and that every single transaction running through your systems is monitored, you've got a great preventative measure.
- It lets people know **that they shouldn't bother**, because they will get caught.

"As fraud schemes become more sophisticated and migratory, access to real time data and the use of advanced data analysis to monitor claims and provider characteristics are critically important." *(Daniel R. Levinson, Inspector General, Office of Inspector General, US Department of Health and Human Services)*

THE POTENTIAL OF DATA ANALYTIC IN FRAUD PREVENTION



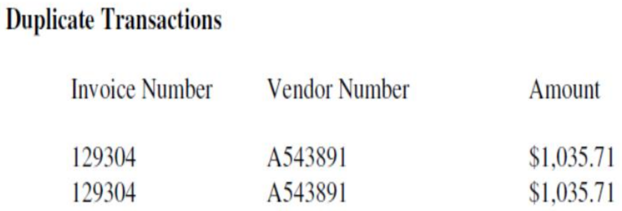
Source: James E. Parry, Predictive Analytics & Fraud Prevention, IBM

DATA ANALYTIC TECHNIQUES

ISACA's 2017
ASIA PACIFIC
CACS

DUPLICATE TRANSACTIONS

1. A simple example of the application of this technique is the search for duplicate transactions, such as the same invoice number - vendor number, payroll credit transaction to same account in a month.
2. Ordinarily, one would expect that invoice number - vendor number combinations, would be unique. Therefore, the existence of transactions with the same invoice number - vendor number combinations would be an unexpected pattern in the data.
3. However, fraud symptoms are only that – symptoms - and care should be taken to properly investigate the transactions before jumping to conclusions.



Duplicate Transactions

Invoice Number	Vendor Number	Amount
129304	A543891	$1,035.71
129304	A543891	$1,035.71

Source: Dave Coderre, ‘The Fraud Toolkit; ‘Fraud Detection: Using Data Analysis Techniques to Detect Fraud’

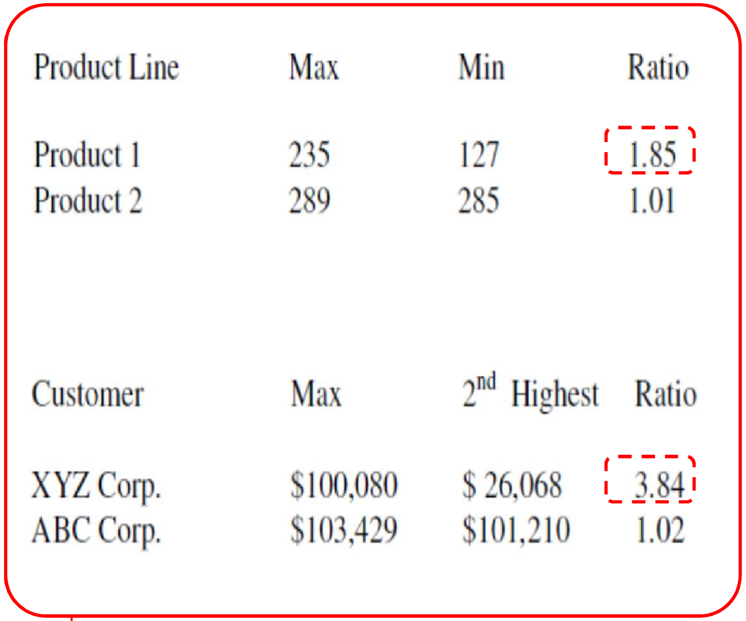
EVEN AMOUNTS

1. Another technique is to identify even amounts, or number that have been rounded up.
2. MOD() function can easily identify these types of even number. I.e. MOD(Amount,100)=0 will identify transactions that are multiple of 100, MOD(Amount,1000)=0 will identify transactions that are multiple of 1000
3. Travel expenses had always be a concern for the auditors as controls were a weak. Employees had a maximum per diem rate when traveling, but had to submit actual receipts to cover the expenses. Another expenses may have their maximums.
4. Some people were charging the maximum rates for meals and hotels even though the receipts did not justify the amounts

Source: Dave Coderre, 'The Fraud Toolkit; 'Fraud Detection: Using Data Analysis Techniques to Detect Fraud'

RATIO ANALYSIS (MINIMUM-MAXIMUM)

1. Like financial ratios that give indications of the relative health of a company, data analysis ratios point to possible symptoms of fraud.
2. The common methods:
   - the ratio of the highest value to the lowest value (Maximum/Minimum)
   - the ratio of the highest value to the next highest (Maximum/2nd Highest)
   - the ratio of the current year to the previous year.
3. If the ratio is close to 1, then they can be sure that there is not much variance between the highest and lowest prices paid. However, if the ratio is large, this could be an indication that too much was paid



Product Line	Max	Min	Ratio
Product 1	235	127	1.85
Product 2	289	285	1.01

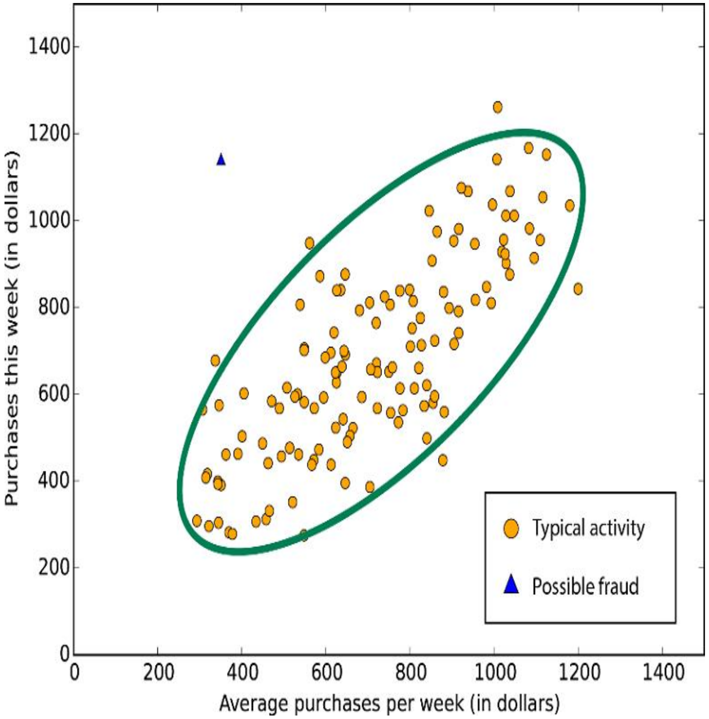
Customer	Max	2nd Highest	Ratio
XYZ Corp.	\$100,080	\$ 26,068	3.84
ABC Corp.	\$103,429	\$101,210	1.02

A large ratio indicates that the Maximum value is significantly larger than the second highest value

Source: Dave Coderre, 'The Fraud Toolkit; 'Fraud Detection: Using Data Analysis Techniques to Detect Fraud'

TREND & REGRESSION ANALYSIS

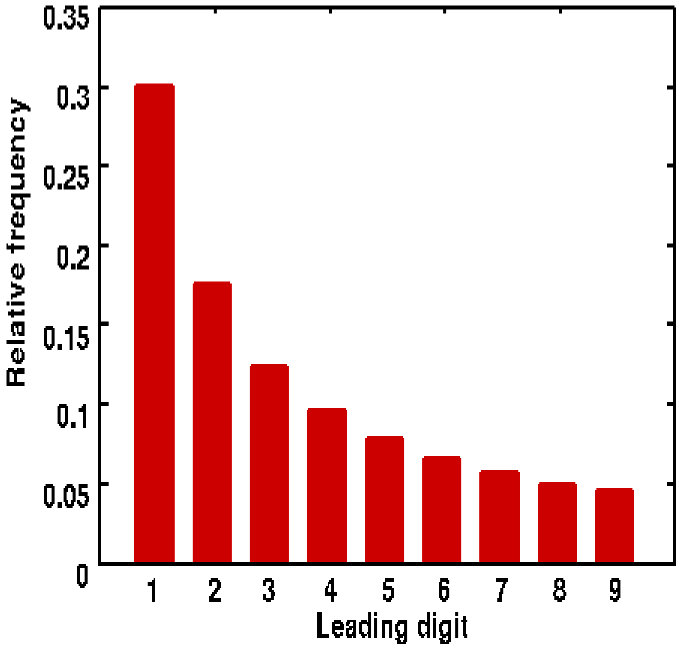
1. Analysis of trends across years, or across departments, divisions, etc. can be very useful in detecting possible frauds.
2. Another useful calculation is the ratio of the current year to the previous year.
3. A high ratio indicates a significant change in the totals. It can be a sudden downturn of upward trend.



Source: Dave Coderre, 'The Fraud Toolkit; 'Fraud Detection: Using Data Analysis Techniques to Detect Fraud'

BENFORD LAW

1. Benford's Law, developed by Frank Benford in the 1920's, makes predictions on the occurrence of digits in the data.
2. Benford's Law concludes that the first digit in a large number of transactions (10,000 plus) will be a '1' more often than a '2'; and a '2' more often than a '3'.
3. Benford calculates that the first digit will be a '1' about 30%, whereas '9' only has an expected frequency of about 5% as the first digit



RULES

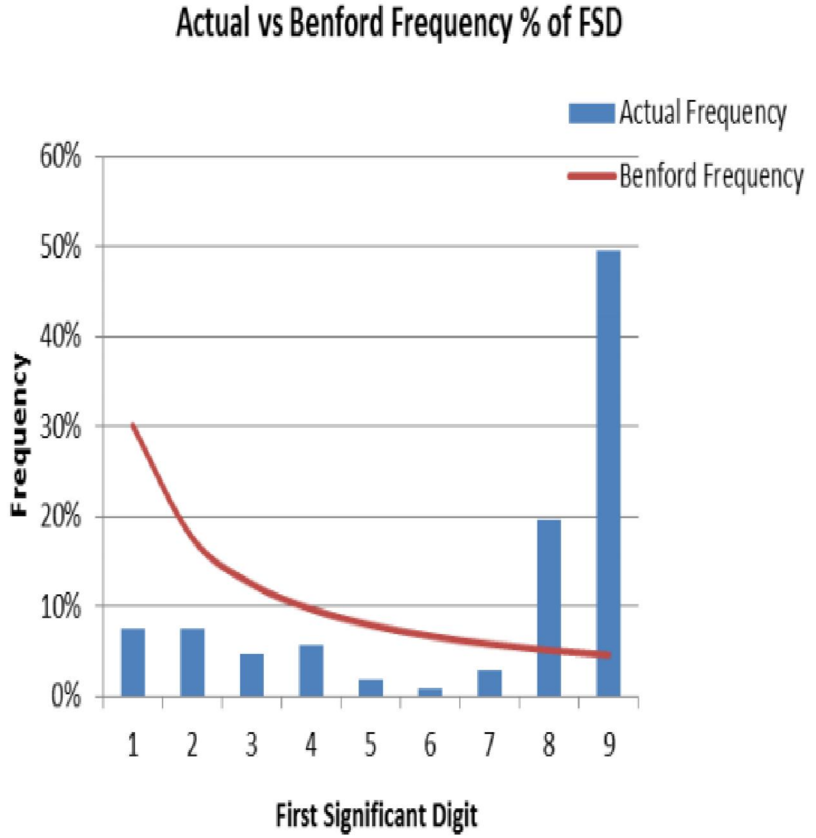
- there should be no set maximum or minimum
- there should be no price break points ($6.12 for all packages under 1 pound, $7.13 for package more than 1 pound and less than 2 pounds)
- numbers should not be assigned, such as policy numbers, social insurance numbers, etc.

Implementation in First Two Digits (FTD):

Expected FTD Frequency = log(1+1/FTD)

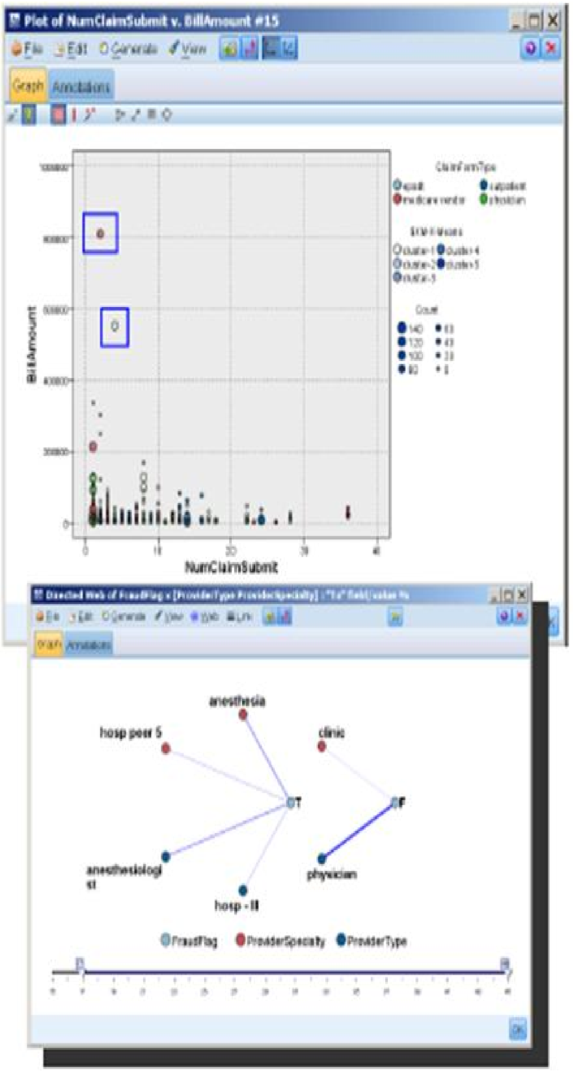
Source: Dave Coderre, 'The Fraud Toolkit; 'Fraud Detection: Using Data Analysis Techniques to Detect Fraud'

BENFORD LAW FOR PROCUREMENT FRAUD



ASSOCIATION, CLUSTERING, & ANOMALY DETECTION

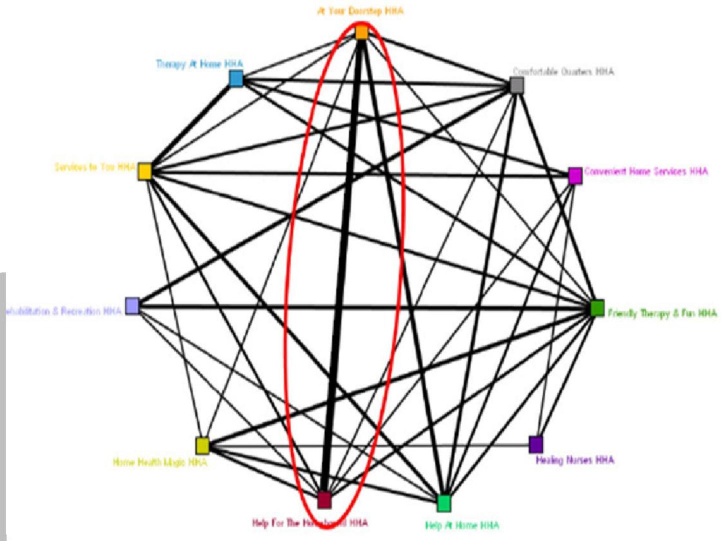
1. Associations:
   - Finds something that occur together (i.e. other events that support a fraudulent event)
   - Associations can exists between any of the attributes
2. Clustering/Anomaly:
   - Reveals natural groups within set of data
   - Encompasses anomaly detection
3. Sequential Associations:
   - Finds association occur in time-oriented data
   - Sequence or order of the events



Source: James E. Parry, Predictive Analytics & Fraud Prevention, IBM

ASSOCIATION, CLUSTERING, & ANOMALY DETECTION

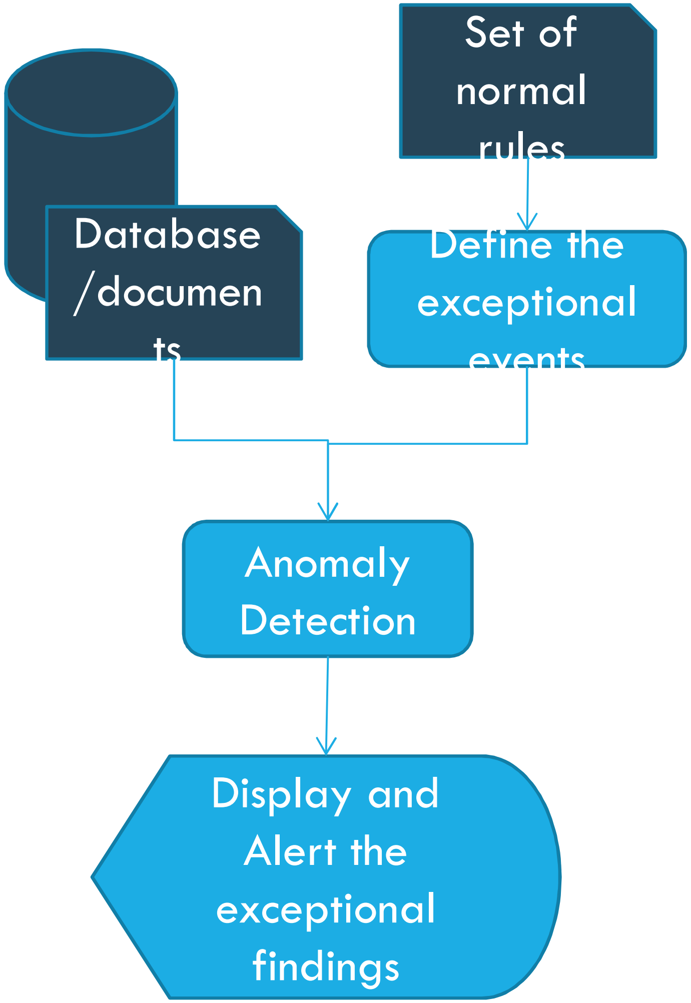
1. Define entities/providers that may shared same attributes
2. Link between each of them by a connecting line of various gradations of width and structure
3. Examples:
   1. Non performing loan with 2 debtors having group relationship.
   2. Payment made to 2/more vendors with same geo-location (distance measurement).
   3. Credit card-holders with same phone-number but different addresses.
   4. Money transfer employee account number



Source: James E. Parry, Predictive Analytics & Fraud Prevention, IBM

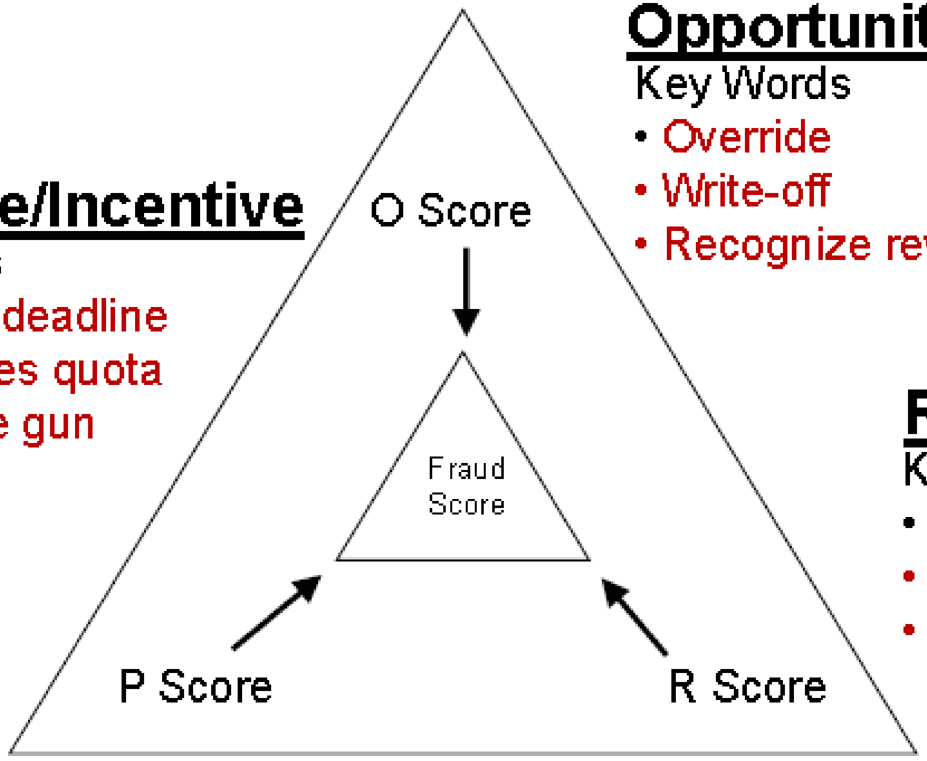
ASSOCIATION, CLUSTERING, & ANOMALY DETECTION

1. Anomaly detection is an exploratory method
2. Designed for real-time and immediate detection of exceptional cases or records that should be included for further analysis
3. Unlike other modeling methods that stores rules about exceptional cases, anomaly detection store information on what normal behavior looks like.
4. Implement as unsupervised learning.


Set of normal rules
Database /documents
Define the exceptional events
Anomaly Detection
Display and Alert the exceptional findings

Source: James E. Parry, Predictive Analytics & Fraud Prevention, IBM

TEXT MINING & FRAUD TRIANGLE



O Score

Fraud Score

P Score

R Score

Pressure/Incentive

Key Words

- Meet the deadline
- Make sales quota
- Under the gun

Opportunity

Key Words

- Override
- Write-off
- Recognize revenue

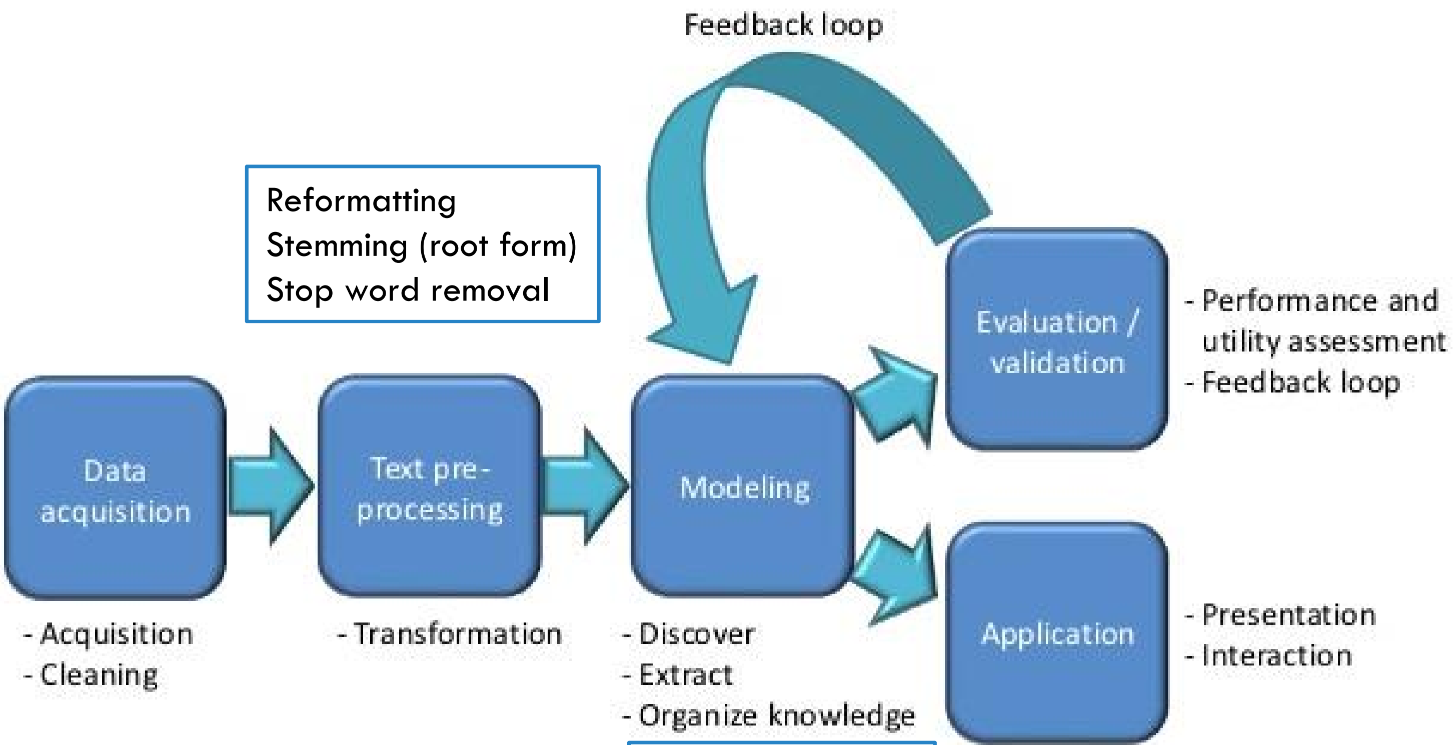
Rationalization

Key Words

- I think it's OK
- Sounds reasonable
- I deserve

Source: Secret Conspiracies & Hidden Patterns: Fraud and Advanced Data Mining, Shauna Woody-Caoussens, Forensic & Valuation Services

TYPICAL TEXT MINING PROCESS



Reformatting
Stemming (root form)
Stop word removal

Simple Counting
Classification
Clustering
Association

Source: Text & text stream mining, Lucca, 2012

FINANCIAL REPORT FRAUD ANALYTIC (MANUFACTURING COMPANIES)

Most non-performing debtors failed to provide a recent updated & audited financial report after their first disbursement

Z-Score = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E

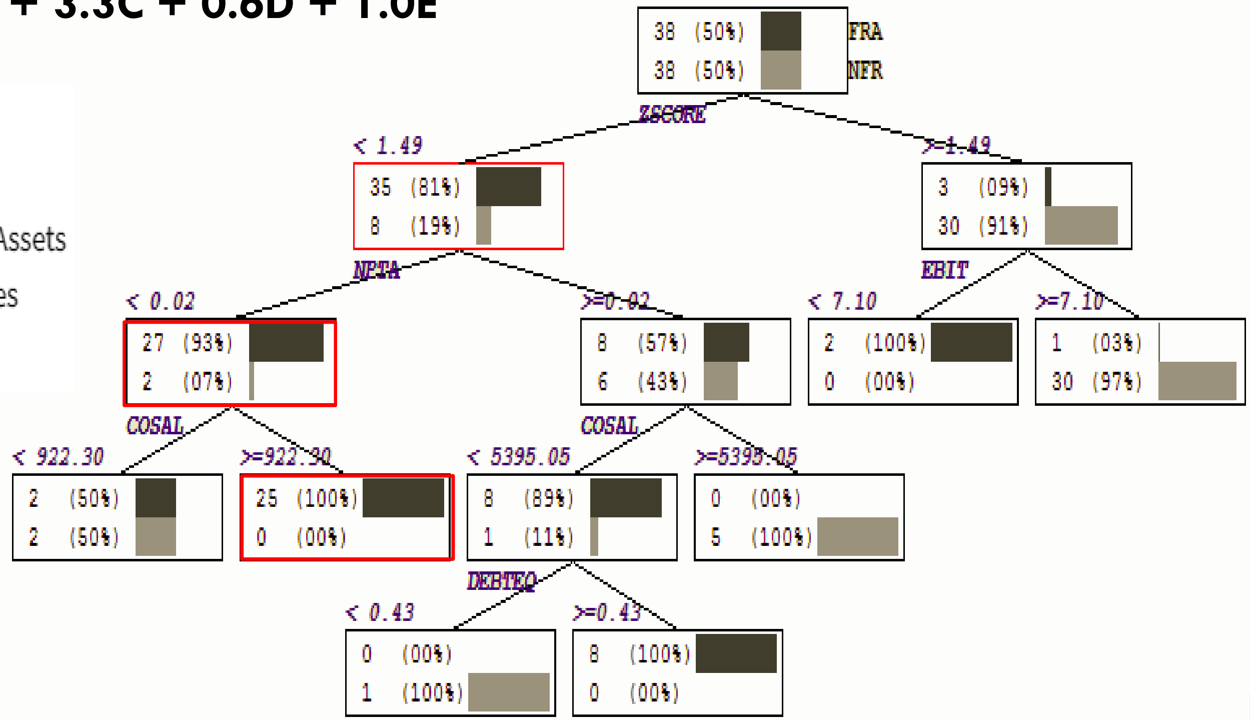
A = Working Capital/Total Assets

B = Retained Earnings/Total Assets

C = Earnings Before Interest & Tax/Total Assets

D = Market Value of Equity/Total Liabilities

E = Sales/Total Assets

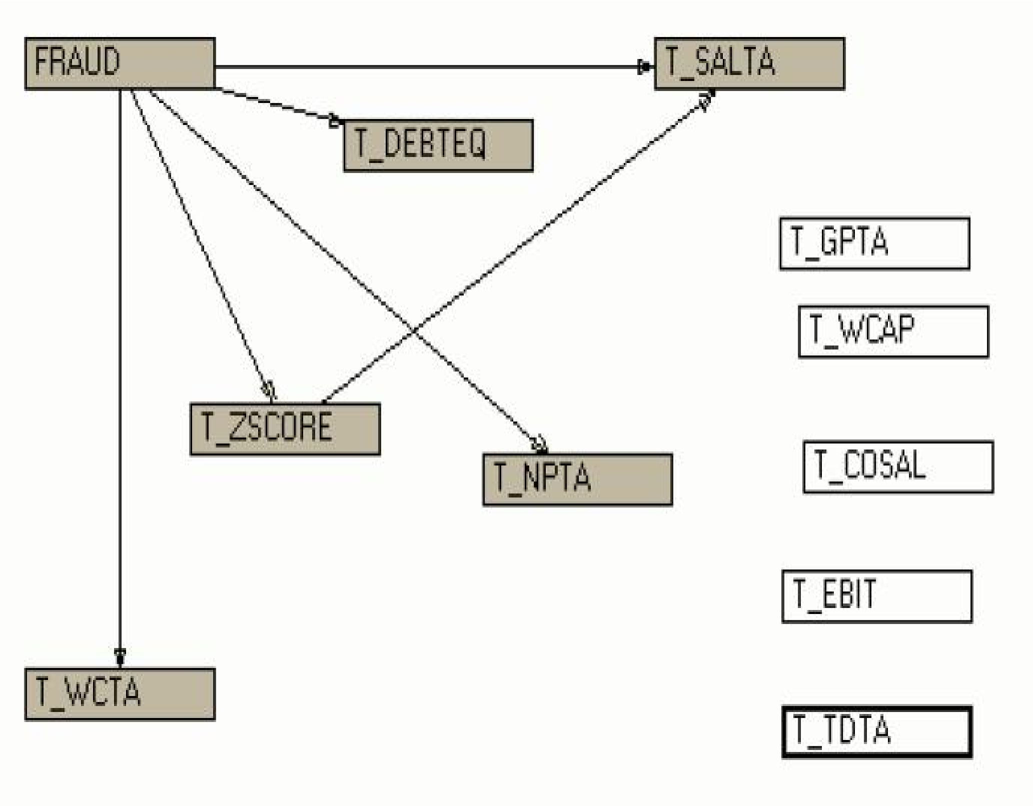


Altman considered a Z score value of 1.81 as a cutoff point to define financial distress for US manufacturing firms, 35 out of 38 fraud companies have low Z-Score <1.49

Source: Detection of Fraudulent Financial Statements through the use of Data Mining Techniques, Efstathios Kirkos, Charalambos Spathis, Yannis Manolopoulos

FINANCIAL REPORT FRAUD ANALYTIC (BAYESIAN BELIEF NETWORK)

According to the network, fraud presents strong dependencies from the input variables ZSCORE, DEBTEQ, NPTA, SALTA and WCTA. Each of these variables expresses a different aspect of a firm's financial status. Z SCORE refers to financial distress, DEBTEQ to leverage; NPTA refers to profitability, SALTA to sales performance and WCTA to solvency.



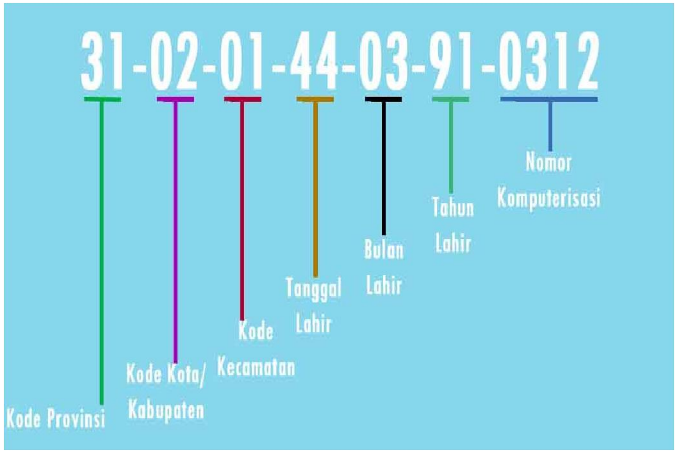
Source: Detection of Fraudulent Financial Statements through the use of Data Mining Techniques, Efstathios Kirkos, Charalambos Spathis, Yannis Manolopoulos

FINANCIAL REPORT FRAUD ANALYTIC (TEXT ANALYSIS/MINING)

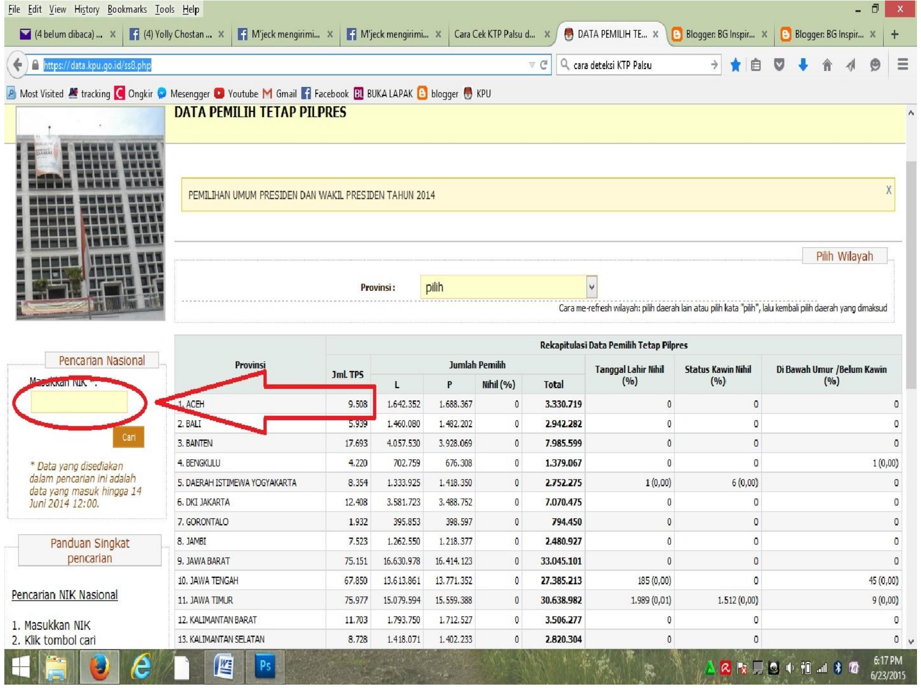
- These disclosures (qualitative narratives) may not contain fraud indicators explicitly; however indicators of fraud can be constructed by **understanding the syntactic as well as semantics of any natural language** because **perpetrators of fraud may camouflage the indicators by using semantic arsenal of the language**.
- In order to conceal the fraudulent activity, perpetrators may use selective sentence constructions, selective adjectives and adverbial phrases.

FAKE ID NUMBER

Indonesian KTP number format



Komisi Pemilihan Umum (Search for KTP)



DATA PEMILIH TETAP PILPRES

PEMILIHAN UMUM PRESIDEN DAN WAKIL PRESIDEN TAHUN 2014

Pilih Wilayah

Provinsi : pilih

Cara me-refresh wilayah: pilih daerah lain atau pilih kata "pilih", lalu kembali pilih daerah yang dimaksud

Pencarian Nasional

Masukkan NIK *:

Cari

Data yang disediakan dalam pencarian ini adalah data yang masuk hingga 14 Juni 2014 12:00.

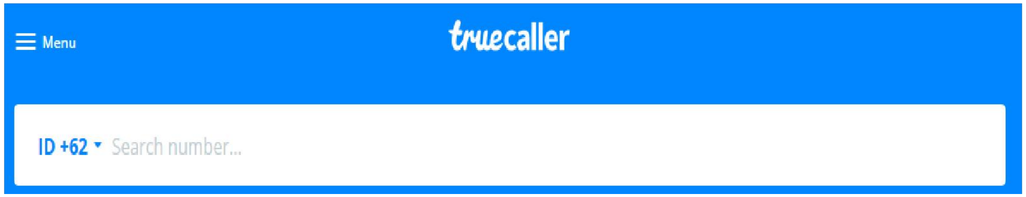
Panduan Singkat pencarian

Pencarian NIK Nasional

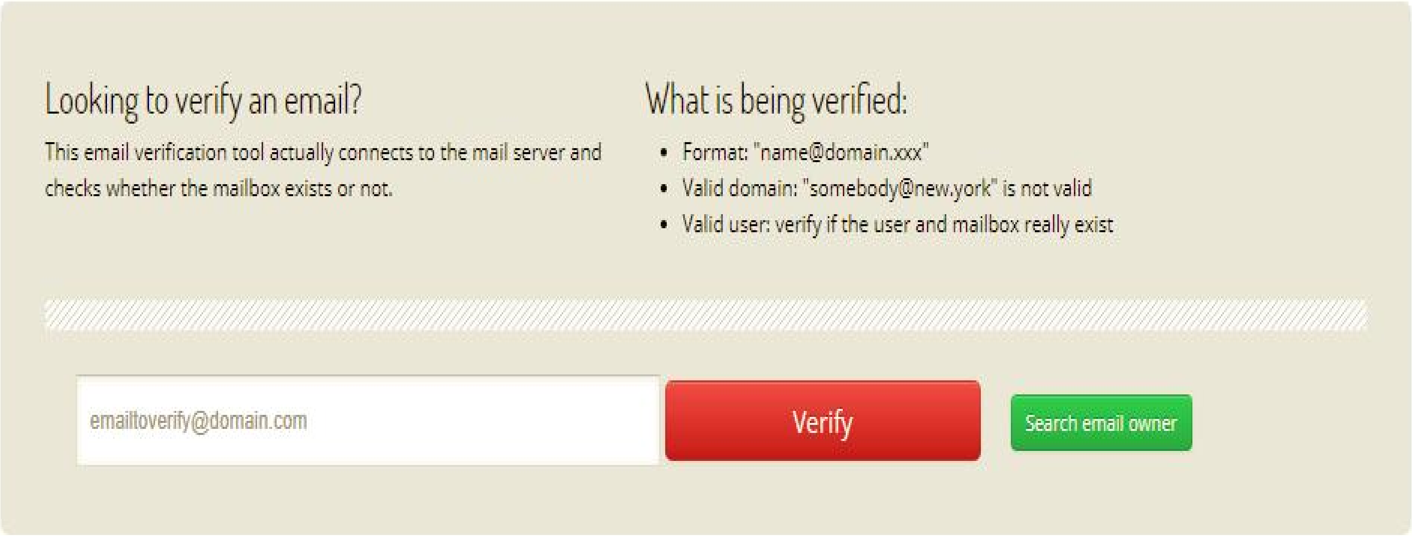
1. Masukkan NIK
2. Klik tombol cari

Provinsi	Rekapitulasi Data Pemilih Tetap Pilpres							
	Jml. TPS	Jumlah Pemilih				Tanggal Lahir Nihil (%)	Status Kawin Nihil (%)	Di Bawah Umur /Belum Kawin (%)
		L	P	Nihil (%)	Total			
1. ACEH	9.508	1.642.352	1.688.367	0	**3.330.719**	0	0	0
2. BALI	5.939	1.460.080	1.482.202	0	**2.942.282**	0	0	0
3. BANTEN	17.693	4.057.530	3.928.069	0	**7.985.599**	0	0	0
4. BENGKULU	4.220	702.759	676.308	0	**1.379.067**	0	0	1 (0,00)
5. DAERAH ISTIMEWA YOGYAKARTA	8.354	1.333.925	1.418.350	0	**2.752.275**	1 (0,00)	6 (0,00)	0
6. DKI JAKARTA	12.408	3.581.723	3.488.752	0	**7.070.475**	0	0	0
7. GORONTALO	1.932	395.853	398.597	0	**794.450**	0	0	0
8. JAMBI	7.523	1.262.550	1.218.377	0	**2.480.927**	0	0	0
9. JAWA BARAT	75.151	16.630.978	16.414.123	0	**33.045.101**	0	0	0
10. JAWA TENGAH	67.850	13.613.861	13.771.352	0	**27.385.213**	185 (0,00)	0	45 (0,00)
11. JAWA TIMUR	75.977	15.079.594	15.559.388	0	**30.638.982**	1.989 (0,01)	1.512 (0,00)	9 (0,00)
12. KALIMANTAN BARAT	11.703	1.793.750	1.712.527	0	**3.506.277**	0	0	0
13. KALIMANTAN SELATAN	8.728	1.418.071	1.402.233	0	**2.820.304**	0	0	0

Caller Name Identification



FAKE ID EMAIL ADDRESS



Looking to verify an email?

This email verification tool actually connects to the mail server and checks whether the mailbox exists or not.

What is being verified:

- Format: "name@domain.xxx"
- Valid domain: "somebody@new.york" is not valid
- Valid user: verify if the user and mailbox really exist

- Use the nslookup –type:mx <email_server>.
- Use telnet command to check the SMTP port and RCPT TO to check the email address.
- Use those commands in VBA Script for Excel formula.

UNSTRUCTURED DATA ANALYSIS - GPS

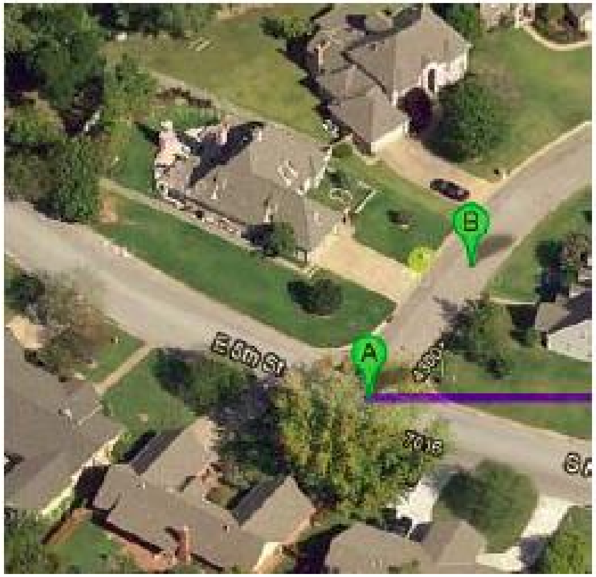
Vendor (A)

Jeremy's Design Company, 123 5th Street, Anytown, MO (Total Payments = $84,337)

Employee (B)

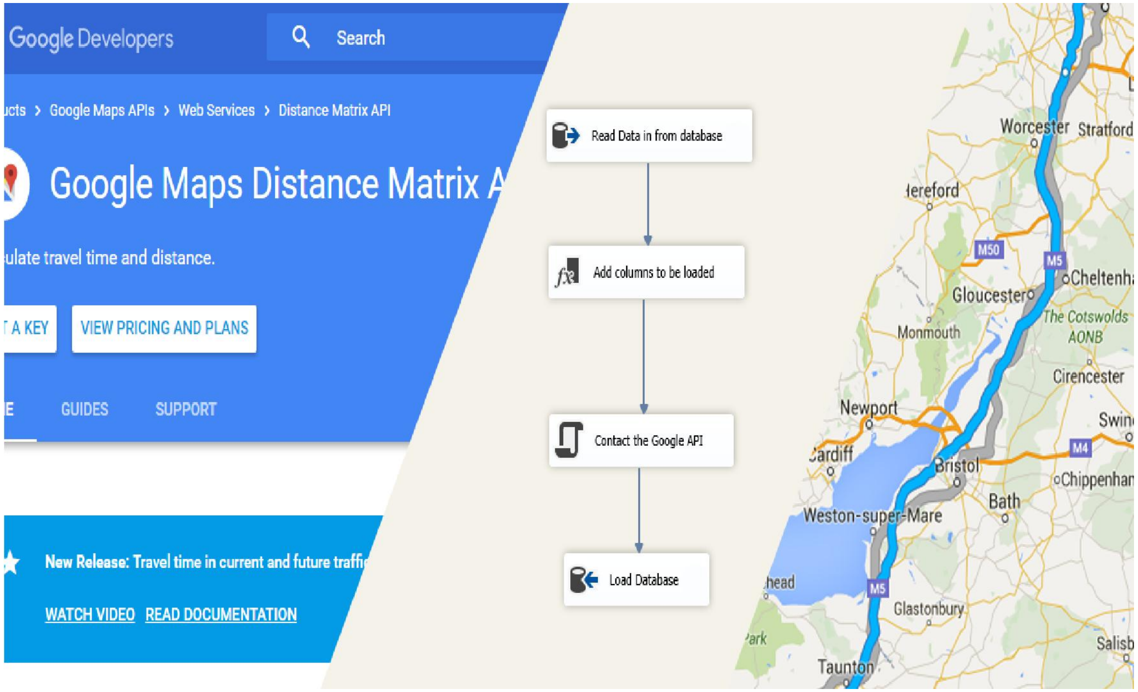
Jeremy Clopton, 4300 Oak Street, Anytown, MO

Only 20% of data in organization is structured data, 80% is unstructured data (not housed in database). However, anti fraud detection is recently focused in the 20%



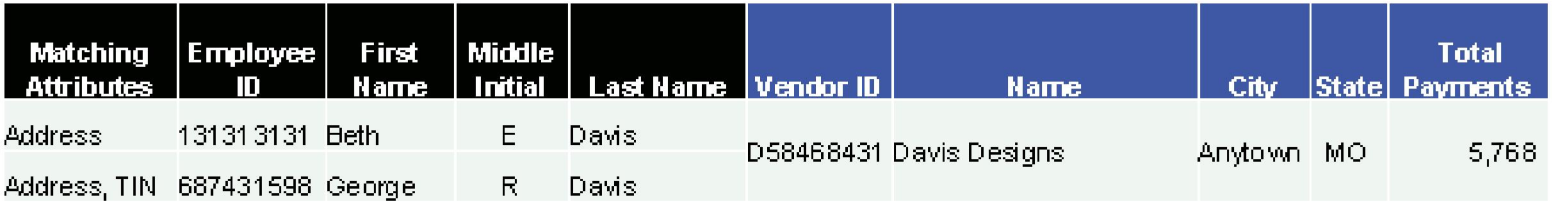
Source: Secret Conspiracies & Hidden Patterns: Fraud and Advanced Data Mining, Shauna Woody-Caoussens, Forensic & Valuation Services

GPS LOCATION TO FIND FAKE ADDRESS

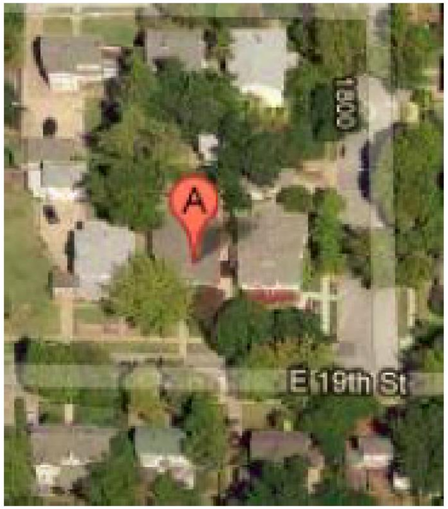


- Use GPS location to find if the address validity
- Use Google Maps Distance Matrix API to get the distance between 2 addresses.
- API function is called from Excel Sheet VBA-Script

VENDOR/EMPLOYEE RELATIONSHIP



Matching Attributes	Employee ID	First Name	Middle Initial	Last Name	Vendor ID	Name	City	State	Total Payments
Address	131313131	Beth	E	Davis	D58468431	Davis Designs	Anytown	MO	5,768
Address, TIN	687431598	George	R	Davis					



Non-Obvious Relationship Association ("Link Analysis")
Linking items that are related but removed by several degrees of separation to mask their relationship.

Latent Semantic Analysis
Concept searching based on tone, recurring themes and communication nuances

Source: Secret Conspiracies & Hidden Patterns: Fraud and Advanced Data Mining, Shauna Woody-Caoussens, Forensic & Valuation Services

THANK YOU