Emerging Technologies and Trends in compliance
Emerging Trends in Compliance

**Advanced Analytics**
- Pre-trade Surveillance
- Machine Learning
- Unstructured Data Analytics

**Model Validation & Audit**
- New Model Validation Methods
- Additional Data Requirements
- Shift in Resource Needs

**AI Process Automation**
- Investigation Optimization
- Periodic Seg. and Tuning
- Regulatory Reporting

**System Requirements**
- Transition to Big Data Infra
- Cloud Computing & Storage
- Innovative Data Encryption

**Emerging Financial Crimes Risks**

**FinTech**
- Virtual Wallets
- Prepaid Cards
- Evolving Products
- Online Gaming
- Electronic Money Transfer

**New Financial Platforms**
- Cryptocurrency
- Blockchain

**New Possibilities from AI and Robotics**
- Increased alert productivity
- Faster processing of alert volumes
- Reduced operational expenses
- Transfer of security risks

Proprietary and Confidential
The Need for Advanced Technologies

- **260 billion**
  - Electronic Payments Global Volume

- **15% to 22%**
  - E-payments Growth

- **$1 million**
  - 1 petabyte Hadoop cluster with 125-250 nodes

- **$4,000 per node**
  - Annual cost of a supported Hadoop distribution

- **$10 to $100 million**
  - Conventional enterprise data warehouse

When asked about their expected use cases and benefits of artificial intelligence in compliance, industry leaders responded:

- **Risk Assessment**
  - Average of 30% to 40% cost savings

- **KYC and AML monitoring**
  - 50% ~ 95% Increase in Efficiency

*Representation proportional to percentage of survey participants with corresponding responses*
# “AI” in Financial Crime

## AI & Machine Learning Techniques

<table>
<thead>
<tr>
<th>Method</th>
<th>Transaction Monitoring</th>
<th>Fraud Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised Clustering</td>
<td>Assist with cross-checking banking data with public records databases for suspicious transactions, and measure the risk associated</td>
<td>Assign a risk score based on known behavior of an individual customer</td>
</tr>
<tr>
<td>Random Forest</td>
<td>Learn transaction behavior for similar customers, and identify outlier transactions and outlier customers</td>
<td>Self-learning mechanism for detecting frauds</td>
</tr>
<tr>
<td>Pattern Recognition</td>
<td></td>
<td>Reduction of false alarms and increases in total savings</td>
</tr>
<tr>
<td>Binary Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typology Based Bayesian Analysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Source** – *Big Shifts in Anti Money Laundering (Booz Allen Hamilton)*

*Machine Learning: Advancing AML technology to identify enterprise risk (ACAMS)*
What is Machine Learning?

*Machine Learning is a discipline of Artificial Intelligence that provides computers with the ability to learn without being explicitly programmed*

According to Tom Mitchell, the former Chair of the Machine Learning Department at Carnegie Mellon University, a machine learning computer program is said to learn from experience (E) with respect to some task (T) and some performance (P), if its performance on (T), as measured by (P), improves with experience (E)\(^1\)

Other Useful “Learning” Terms

**Data science** is an interdisciplinary field about processes and systems to extract knowledge or insights from data.

**Artificial Intelligence (AI)** is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence.¹

**Machine learning** is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed.

**Deep learning** is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations.

**Statistical model** is a class of mathematical model, which embodies a set of assumptions concerning the generation of some sample data, and similar data from a larger population. A statistical model represents, often in considerably idealized form, the data-generating process.
Types of Machine Learning Problems

There are three types of machine learning problems: supervised, unsupervised, and reinforcement learning; the team will focus on supervised and unsupervised learning throughout the workshop.
Link Analysis
The Panama Papers are a set of 11.5 million documents and database records. They detail the relationships between 214,000+ offshore companies and individuals. The documents were leaked from Mossack Fonseca to a German newspaper, which analysed the data alongside the ICIJ*. On 9th May 2016, the ICIJ released data that named the companies, individuals and relationships in the Mossack Fonseca database.

*About the Panama Papers,* Süddeutsche Zeitung, http://panamapapers.sueddeutsche.de/articles/56febff0a1bb8d3c3495adf4/

---

**The scale of the leak**

<table>
<thead>
<tr>
<th>Volume of data compared to previous leaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7 GB Cablegate/Wikileaks (2010)</td>
</tr>
<tr>
<td>260 GB Offshore-Leaks/ICIJ (2013)</td>
</tr>
<tr>
<td>4 GB Luxembourg Leaks/ICIJ (2014)</td>
</tr>
<tr>
<td>3.3 GB Swiss Leaks/ICIJ (2015)</td>
</tr>
<tr>
<td>≥ 2.6 TB Panama Papers/ICIJ (2016)</td>
</tr>
</tbody>
</table>
Applying Link Analysis to the Panama Papers

- Previous methods of analysing risk were limited to direct matches to sanctioned entities and manual investigation.
- Using a sophisticated risk scoring algorithm, it’s possible to assign a propagated risk score to every connected entity in the network via shockwave risk rating.
- The risk level spreads outwards from each high risk entity, building a complete risk landscape across the data.

1. Locate all high risk entities in the network, such as sanctioned entities and PEPs.
2. Propagate risk throughout the network.
3. Assign a score to each entity along the path, reducing the strength of the shockwave at each step, according to configurable settings.
Applying Link Analysis to the Panama Papers

1. First, customer data is loaded alongside the other data sources

2. Our algorithms then match similar names to each other

3. A set of known high risk entities (e.g., PEPs and press report mentions) are built into the enriched database

4. The “risk shockwaves” propagate through the enriched network of data, producing a risk score for every entity in the path

5. ABC Limited is marked with a risk score because of its connections to other entities, allowing us to flag it for further investigation
**Applying Link Analysis to the Panama Papers**

- When the screening processing is complete, the results are displayed in an interactive, standalone report file.
- A report is generated for each matching screened entity, showing their level of risk and connections.
- This allows an investigator to determine whether the risk is genuine and could represent a damaging relationship.
- For large numbers of identified entities, a high-volume reporting and review platform is an alternative solution.
Combining information with your data model

1. Relationship information in the ICIJ Offshore Leaks, Panama Papers and Bahamas Papers data

2. Relationship information from new leaked data sources

3. Entity-level data from third party data sources

An enriched database of relationships and entity details

Stanley Example

- Historical OFAC list
- Politically Exposed Person

- Registered address: 123 Greenland Park Gardens, London SE1 2AD
- OFAC active dates: April 2007 to present
- PEP category: INDIVIDUAL
- PEP exposure index: 99

...
Recent Trends in AI & Robotics as it Relates to Data

Consumer Facing Robots
Autonomous robots such as NAO & Pepper and Nina Web Assistant designed to interact with customers and answer simple questions

Cost Reduction with Artificial Intelligence
15% cost savings across front-office, middle-office and back-office functions

Changing Systems & Processes

<table>
<thead>
<tr>
<th>Today’s systems &amp; processes</th>
<th>Tomorrow’s systems &amp; processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differing standards on data elements</td>
<td>Standardized data models &amp; enhancement methods</td>
</tr>
<tr>
<td>Limited to no external data utilized</td>
<td>Ingestion of external data</td>
</tr>
<tr>
<td>Limited scalability in systems /databases</td>
<td>Dynamic and intelligent tests for data quality</td>
</tr>
<tr>
<td>Varied data enhancement methods</td>
<td>Easily scalable model for systems and tools</td>
</tr>
</tbody>
</table>

Impacts and Potential Pitfalls

- **Increased Alert Productivity**: A more efficient approach to understanding alert to SAR ratios based on statistical principles
- **Reduced operational expenses**: Increased alert productivity reduces the need to support high investigation volumes
- **Improved market transparency**: Scalable financial crime data environments allow for easily distributed
- **System Failure**: A single system failure can potentially affect all parts of the connected process
- **System Complexity**: Incorporating relationships between all financial crime systems potentially increases system complexity

Proprietary and Confidential
Robotic Process Automation (RPA) vs. Traditional Automation

The concept of automating tasks is not new, but traditional automation methods are evolving using Robotic Process Automation (RPA) platforms that mimic the individual actions and process steps that are made by humans; taking automation to another level outside of just automation within individual applications.

**Human Element**
RPA has the potential to mimic behaviour of human across tasks and platforms

**Ease of Implementation**
Implementation is faster and easier than traditional automation methods

**Adaptability**
RPA can easily be altered since it is not dependent on backend settings

**Similarities**
- Streamline Process
- Increase Efficiency
- Lower Overhead
- Reduce Lead Time

**Structural Differences**
Traditional automation changes the underlying back-end functioning of the system

**Self-Contained Processes**
Efficient high-frequency processes are ideal candidates for traditional automation

**Adaptability**
Traditional automation relies on rules and scripts making it less adaptable
Robotic Process Automation (RPA)

Software robots can replicate human action by interacting with desktop applications and processing transactions in the same way the human employee would, reducing overhead, lead time and errors.

**RPA Benefits**

- **At least 30%** of processes could be automated for higher efficiency gains.
- **Cost reduction of up to 50%**
- **RPA works 4-5X faster than employees and 24/7**
- Humans typically make 10 errors during a 100-step process, robots are 100% accurate.

**RPA Adoption**

- Market to reach **$4.98 BN** globally by 2020.
- **50% plan to or are actively pursuing RPA pilots**
- **40% believe RPA tools are the most enabling technology today**
- **28% of companies have implemented RPA**
**Types of Robotics**

Robotics are generally comprised of RPA and IPA technologies, which are described below.

Robotics is not traditionally a part of a company’s information technology infrastructure, but rather sits on top of it.

<table>
<thead>
<tr>
<th>Known as</th>
<th>Why are they used?</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Robotics Process Automation” or RPA</td>
<td>Replaces humans <em>who plug the gaps</em> between non STP systems — e.g., &quot;low value activity&quot;</td>
<td>Logic Driven, software executes pre-programmed tasks as defined by the robot designer</td>
</tr>
<tr>
<td>“Intelligent Process Automation” or IPA</td>
<td>Replaces human reasoning (complex, non-linear)</td>
<td>Logic Driven + “machine learning”. Software “learns” trends in data and uses statistics to execute tasks based on learned trends.</td>
</tr>
</tbody>
</table>

**How is RPA different than traditional application “automation”?**

- RPA software interacts with presentation layer
- Traditional application automation interacts with business logic and data access layers
**Enterprise Adoption of Robotics**

Workforce efficiency and effectiveness have long been a top enterprise priority. As technology continues to rapidly evolve in the space, decades of proven labor arbitrage and shared services are no longer the most constructive approach.

**Evolution of Robotics**

- **Human Workforce**: Presence
- **Use of Robotics**: Migration of manual tasks to offshore locations – effective strategy for decades, expanding to more sophisticated tasks

  - Efficiency vs. Effectiveness
  - Labor arbitrage savings 20-30%
  - Lean principles drive additional 5-15% efficiency

- **Mainstream Adoption**: 2005-2015
  - Location Strategy, Outsourcing & Lean
  - Apps to create “bots” to replace humans who plug the gaps between non STP systems – “low value activity”
  - Logic driven, software executes pre-programmed tasks as defined by the robot designer
  - For certain activities - estimated 3 times savings over traditional offshoring / BPO

- **Fast Emerging**: 2015-2017
  - Robotics Process Automation (RPA)
  - Intelligent Process Automation (IPA)
  - Incorporates cognitive intelligence to execute tasks and update rules based on “learned” trends, requiring minimal human oversight
  - Estimated to replace 110-140M knowledge workers by 2025
  - Financial Services, Health and High Tech are emerging industry leaders

---

Robotic Process Automation (RPA) and Intelligent Process Automation (IPA)

Robotic Process Automation (RPA) and Intelligent Process Automation (IPA) are new tools within the overall set of automation technologies.

<table>
<thead>
<tr>
<th>Description</th>
<th>Automation Sophistication</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule-based automation within specified software such as MS Excel or web-based.</td>
<td>Robotic Process Automation (RPA) – or Robotics Desktop Automation (RDA)</td>
<td>Computer generated virtual assistant that simulates a conversation to provide guidance</td>
</tr>
<tr>
<td>Using software applications and integrating those including restructuring labour, in order to minimise costs.</td>
<td>Natural Language Processing (NLP)</td>
<td>Use of a cognitive framework to develop insights, applying learning from experience to data from multiple sources</td>
</tr>
<tr>
<td>Activities that are repeatable within a specific application (e.g., Excel), providing the user with a way to automate a repeatable process with highly structured data.</td>
<td>Cognitive Computing (IPA)</td>
<td></td>
</tr>
<tr>
<td>Processes that require to use multiple systems together with specified logic e.g. providing customer service by different teams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour-intensive repetitive activities that need significant amount of data processing across multiple applications will benefit through RPA without the need to change existing systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A digital assistant removes the need to have manual conversations and responses such as offered within customer service function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Artificial Intelligence is increasing in areas where vast array of data processing is required to make decision while considering the overall context</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proprietary and Confidential
Use Case 1: KYC Process Automation

Gathering and updating customer information is a typical, repetitive operation involving manual verifications and updating of multiple systems in which accuracy is critical to business functions. Processes such as these are ideal for automation.

Customer Data Collection

Customer updates information (name, address, phone, dependencies, etc.)

Updated information is placed in a work item which is queued for processing

Automated Processing of Customer Data

ID verification is performed on updated customer information

Back office system(s) need to be updated

Outcome successful or further information request is sent to the customer

RPA System

Case Study Key Outcomes*

- 570 cases from live queue processed
- On average, robots were three times faster than human operator
- Robots were 100% accurate
- Successful integration of RPA with existing IT
- 51% of work items fully completed by the automation robots

Source – Automate and Mundane for Accelerated Cost Reduction (NICE)
**Use Case 2: Automated Alert Investigation**

Consolidation of disparate data sources and the automation of alert investigations enables L1 reviewers and FIUs to decision on alerts faster and more accurately; reducing operational overhead and process lead time.

### Automated Aggregation of Customer Data

- Disparate customer data sources utilized by investigators are automatically aggregated and linked to form an AML single customer view.

### Automated Alert Investigation and Adjudication – Unsupervised Learning

- Alerts generated and matched against customer data.
- External sources are pulled.
- L1 review is automated using rules based on historical investigator decisions and continuous unsupervised learning.
- Process is time and labor efficient and reduces overall program costs.

### Current Manual Investigation Process

- Alerts are generated across each system individually and not linked to specific customers across FIUs.
- L1 review performed by investigators who sort through and collect data across systems, LOB contacts and the web. Data must also be compiled and verified.
- Investigators must manually sift through data and subjectively review and adjudicate alerts.
- Process is time and labor intensive and requires a larger headcount to support.

AML Systems

- Escalation
- False-Positive

EDW

- False-Positive

AML Systems

- Escalation
**Use Case 3: Automated data gathering and collection**

Automating data gathering and consolidation represents a significant opportunity to streamline the investigative process due to the sheer number of inputs that can flow into a case management system.

### Inputs

**Customer information**
- Customer onboarding
- Customer due diligence and periodic review
- Customer document repositories
- Credit reports

**Surveillance system alerts**
- Transaction monitoring system
- Trade surveillance system
- Cybersecurity system
- Other detection systems

**List screening/filtering**
- Office of Foreign Assets Control (OFAC) lists
- Politically Exposed Persons (PEP) lists
- Internal watch lists
- Negative news lists
- Beneficial ownership information

**Public sources**
- Regulations and regulatory updates
- Web searches
- Adverse media citations

**Internal sources**
- Customer Relationship Management system
- Internal policies and procedures
- Analytic/Detection system
- Internally collected notes and analyses

### Outputs

- Regulatory filings
- Feedback to detection or monitoring systems
- Feedback to other enterprise systems
Alert routing and alert nudging to improve operational efficiency...

Based on a number of factors, the system can automatically route alerts to an appropriate analyst, which can save time and help reduce risk. The system can then suggest a series of next steps based on the particular attributes and risk score of the alert.

<table>
<thead>
<tr>
<th>Alert number</th>
<th>Scenario</th>
<th>Account No.</th>
<th>Transaction count</th>
<th>Customer type</th>
<th>Total transaction amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML4952</td>
<td>Sequentially Numbered Checks &amp; MI</td>
<td>AC000001</td>
<td>12</td>
<td>Non Profit</td>
<td>$2,000</td>
</tr>
</tbody>
</table>

Joe Sample is the first choice

If Joe Sample is not available:

1. Joe Sample 92%
2. Ed Person 87%
3. Sue Jones 65%

Investigator concludes that behavior is suspicious, system recommends escalation to specific team leads

System identifies aberrant activity, prompts investigator to request information from client's relationship manager

Investigator flags interesting transactions, system adds joint account ownership information

Investigator flags relevant counterparties, system adds relevant internal/external information

System shows that client has significant news presence

System presents investigator with client's transaction history

Alert is automatically routed to experienced investigator
Automated alert risk scoring

Automated alert risk scoring can optimize the investigation process while also ensuring that higher-risk alerts are prioritized. Risk scoring identifies risk factors (such as count, account age, historical profile, transaction geography, etc.) that can predict the productivity of an alert.

Statistical models and business rules score and prioritize alerts based on risk level.

Machine learning allows rules to adapt to changes in alert characteristics and trends.

Alerts Generated, Scored and Matched Against Aggregated Customer Data

AML Analytics Environment

9

6

2

Sanctions, PEP and watchlist screening

Transaction monitoring systems and models

Capital markets, wealth and asset management, private banking and trade finance due diligence
Use Case 4: Data Quality Analysis Automation

A data quality assessment involves performing various data quality validations to assess the potential impact of data quality issues which is critical in determining the effectiveness of the AML monitoring program. Repetitive tasks such as processing of source system data can be automated to gain efficiency and accuracy.

Source Data Feeds

Data Warehouse – Source System Destination

Source System A

Source System B

Automated Data Quality Checks and Processing

Profile, Distribution and Pattern Analysis is automatically performed on data feed

KDEs are identified and issue records are documented

Immediate response team is contacted to address KDE issue records

Non-KDEs are identified and issue records are documented

Issue records are sent to source system owners to be addressed

RPA System

Update Issue Records

Issue records are then fixed at the source system by appropriate source system owners for re-processing

Machine learning can be applied to automate future corrections

Repetitive tasks such as processing of source system data can be automated to gain efficiency and accuracy.
Use Case 5: Automated Model and Scenario Tuning

The tuning of model and scenario parameters and thresholds is a critical component that defines an effective and efficient AML program. The tuning process is laborious and time consuming, but can be automated to assess and optimize model parameter and threshold values based on unsupervised learning procedures.

Transaction Monitoring Systems generate alerts based on business rules that utilize specific parameter values and thresholds, which are criteria for what qualifies as an alert.

These parameter and threshold values are tuned following a predominantly manual process performed by teams of analysts.

The tuning process can be automated across all stages:
- Baseline Thresholds
- Sample Sizes and ATL/BTL Sampling
- Parameter Analysis
- Threshold Assessment

Operational efficiency:
- Consistent, streamlined processes and automation
- Rapid changes to patterns in alerts based on unstructured learning outcomes
- Reduction in overhead and process lead time

Based on the outcome of automated analyses and sampling results, model settings are then automatically modified.

Note: The sampling process still requires investigator feedback and decisioning.
## Considerations: What we have heard from our clients

<table>
<thead>
<tr>
<th>Consideration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology / Business Partnership</td>
<td>• The right level of partnership and ownership between <strong>IT &amp; business areas</strong></td>
</tr>
</tbody>
</table>
| Governance                        | • **Governance structure** (e.g., federated vs. centralized).  
• Where do **skillsets need to be developed**.                                                                                                    |
| Education                         | • Understanding **what robotics is** and the language used. (e.g., IPA, RPA, Digital Labor, Robotics, Machine Learning, etc.)  
• Making sure **staff understand capabilities and purpose** of digital labor programs – it isn’t about taking jobs, it is about making staff more effective. |
| Understanding the Hype             | • **Realistic value** that robotics can bring to the organization.                                                                               |
| Risk                              | • **Regulator response / considerations** in how to deal with the new automation tools.  
• Getting internal control groups **comfortable with the use of robotics** and policies and procedures governing their use.  
• Creating and agreeing **security and entitlement policies**                                                                                     |
| Organizational Responsibility     | • Having a clear understanding who is **responsible for the robot**, how the robot **works with / interacts with humans** and who gets the “call” when the robot malfunctions. |
| Support Structure                 | • How does change management work and what is the process for reprogramming the robot when these changes occur?                            |
| Sourcing                          | • How does robotics impact relationship / cost structure with business process outsourcers (BPOs) – is there opportunity to insource or re-negotiate contracts? |
**Commonly Used RPA Vendors**

The following list contains a number of Robotic Process Automation vendors, but is not exhaustive. Each vendor provides elements of RPA functionality, but are differentiated through how processes are established, additional machine learning features, and data interaction, among others.
## RPA Vendor Landscape

### NICE - Robotic Automation

**Strengths:**
- **Ideal for large sets of structured data** Nice Robotic Automation automates data processing, specifically back-office finance and administrative processes by mimicking human interface interactions and through machine learning.
- **Effortless data process optimization** through integration that leverages machine learning capabilities to optimize analytics tools.
- **Improves processing time, data organization and reliability while maintaining ease of scalability** while also offering customer support to navigate complex data.

**Weaknesses:**
- **Unstructured data** may hinder the efficacy of the tool’s performance.

### Work Fusion

**Strengths:**
- **Quality integration of security and compliance processes** into model.
- **Advanced rule-based feature set allows for efficient dealings with unstructured, semi-structured and structured data sets**.
- **Advanced performance regarding IT system gap filling and simplistic process implementation**.

**Weaknesses:**
- **No machine learning features** or action recordings as of now.

### Automation Anywhere

**Strengths:**
- **Easily accessible and customizable features** allow for simple integration into preexisting tools and functionalities.
- **Automation Anywhere gives users the power to automate tasks on a user-friendly interface** that makes process automation intuitive to a user with minimal technical background. Mimics user actions and process flow.
- **Drives more effective decision-making to ultimately achieve a zero error rate** through creating concrete guidelines and error reporting that enables continuous correction of data processes.

**Weaknesses:**
- **No machine learning capabilities** so processes must be updated by users every time the data format or process is challenged or modified.

### UiPath

**Strengths:**
- **Simplistic tool creation and modification** to ensure business needs are met in a seamless and timely manner.
- **Extensive customer support** allows for ease of implementation and troubleshooting, as well as on-the-go tool-related consulting services for clients.
- **Executes line of business forms and data management operations with machine learning processes that require minimal user upkeep**.

**Weaknesses:**
- **Machine learning command center capabilities are relatively limited** despite high performance of available features.

Source: vendor materials. Information has not been validated by PwC and does not reflect PwC views on the product, service or vendor.

Proprietary and Confidential
Automated link and pattern analysis

As data gets loaded into the case management system, it can be automatically tagged to an investigation, enabling identification of connections between relevant data points.

- **Bad actor hacks into a checking account (cybersecurity event)**

- **Same bad actor attempts a wire transfer (fraud event)**

- **Neither event triggers a money-laundering alert, but a party related to the bad actor is under investigation**

System displays connections through visual link analysis

Automation helps connect the dots

Source: PwC
**Automated link and pattern analysis**

Historical alerts can provide context for new alerts. The system provides that context by considering each alert as a series of attributes, which it weighs according to various risk factors.

<table>
<thead>
<tr>
<th>Current-alert attribute</th>
<th>Current-alert attribute value</th>
<th>Attribute weight</th>
<th>% match with historical alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account number</td>
<td>AC 111002</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Customer name</td>
<td>Ned A. Sample</td>
<td>12%</td>
<td>0%</td>
</tr>
<tr>
<td>Customer type</td>
<td>Retail</td>
<td>25%</td>
<td>100%</td>
</tr>
<tr>
<td>Customer business</td>
<td>Attorney</td>
<td>4%</td>
<td>100%</td>
</tr>
<tr>
<td>Transaction type</td>
<td>Wire</td>
<td>12%</td>
<td>100%</td>
</tr>
<tr>
<td>Transaction amount</td>
<td>$2,500</td>
<td>8%</td>
<td>95%</td>
</tr>
<tr>
<td>Intermediary bank</td>
<td>Generic National Bank</td>
<td>4%</td>
<td>100%</td>
</tr>
<tr>
<td>Beneficiary bank</td>
<td>Big City Bancorp NAM</td>
<td>7%</td>
<td>40%</td>
</tr>
<tr>
<td>Beneficiary name</td>
<td>John Doe</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Beneficiary country of residence</td>
<td>Belarus</td>
<td>12%</td>
<td>75%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>100%</strong></td>
<td><strong>83.6%</strong></td>
</tr>
</tbody>
</table>
Detecting Unauthorized Trading Activity

PwC developed an unsupervised approach to quickly reduce the population of customers who may have been impacted by the unauthorized trading scheme through intelligent feature engineering and the k-means clustering algorithm; team was able to quickly isolate 42 customers for detailed review out of population of greater than 1300

**Transformation of Data**
- The input data needed to be transformed to adjust for metrics that have highly skewed distributions\(^1\)
- Skewed distributions were transformed by taking the natural logarithm (logarithm to the base of e)
- All metrics are further normalized by subtracting the variable (metric) mean and dividing by the standard deviation of the variable
- 17 metrics in total were included as parameters in the k-means clustering analysis (15 transactional, 2 financial)

**Identification of Number of Clusters**
- Tightness of clusters are analyzed to determine optimal number of clusters for the analysis given the customer population
- The number of clusters was determined to be five as the best representation of the ‘elbow’ of the graph

**K-means Clustering Results**
- Clusters are formed based on the similarity of all 17 input parameters; forming 6 natural groupings
- Cluster centroids\(^2\) are shown as a simplified, 2D illustration of the clusters relative differences

---

1. Avg. amt./mo., deposits, and withdrawals were all transformed with natural logarithm
2. Centroids are averages across ‘centers’ of all parameters in a cluster

Proprietary and Confidential
Anomaly Detection In Retail Banking

PwC performed a significant amount of feature engineering, customer segmentation and advanced analytical analysis to develop the two algorithms to determine anomalies in transactions.

PwC performed feature generation based on customer attributes...

...and community detection based on k Nearest Neighbors to identify distinct customer segments through transaction profiles...

...and finally, identified anomalous activity based on isolation forest and path to anomaly

...then, learnt dynamic Bayesian network and identified anomalous activity based on predicted probabilities and anomaly index...
Peter Golden

**Background**

- Peter I. Golden is a member of PwC’s Financial Crime Unit Technology practice working with financial services clients in the area of financial crimes risk and compliance.
- He has worked extensively with Anti-Money Laundering (AML) solutions, regulatory reporting requirements, IT Governance and Processes, and risk management (Operational, Market, Credit, Liquidity) systems.
- Pete has assisted clients in designing and implementing new visual and predictive analytics solutions, assessing and remediating data quality issues and enhancing institution’s technical capabilities around data. Pete has a strong technical background in programming languages, traditional and non-traditional data storage technologies, visual analytics, and predictive analytics.
- Prior to PwC, Pete lead the North American risk technology team at Goldman Sachs Internal Audit. In his role, Pete lead the testing of the design and effectiveness of technology controls, including data governance, technology policies and procedures, and the implementation of automated controls. Pete started his career as a Nuclear submarine officer in the United States Navy.
- Pete holds a Master of Science in Predictive Analytics from Northwestern University as well as a Bachelors of Science in Engineering (Computer Science, Electrical Engineering, and Computer Engineering) from Duke University. Pete also holds the Certified Anti-Money Laundering Specialist (CAMS) designation.

**Relevant experience**

- For a top 10 global investment bank regulated by the Monetary Authority of Singapore (MAS), led data mining efforts to verify data integrity of transaction data used for AML and transaction monitoring systems, including mapping out data flows from front office client account and transaction systems, compliance monitoring and screening systems, and treasury payment systems. Analytical activities included SQL scripting, data extraction, transformation and loading, fuzzy matching analysis and procedural development to automate data testing.
- For a top 10 global consumer bank, led assessment of data quality and controls for global strategic KYC system. Project included inventory of global data feeds, assessment of in place controls, and source to surveillance data quality reconciliation of critical data elements in customer data flowing into KYC system.
- For a major retail bank, directed the successful rollout of project and portfolio management analytic and reporting tools supporting $1.3 billion of technology investment, including integration with existing data repositories, security implementation, and performance optimization; three previous attempts by other parties had failed; project delivered under budget.
Sang Chung

Background

Sang has over 12 years of experience in the financial services industry, mainly in an IT advisory capacity in the Anti Money Laundering domain. Sang has extensive experience working with the compliance groups of financial institutions to design, implement and optimize the transaction monitoring systems for various lines of business leveraging big data tools and technologies. Sang also has extensive experiences in compliance softwares including Oracle FCCM(Mantas), Actimize SAM and ERCM and Custom solutions.

Sang has led AML-TMS enhancement/remediation efforts for a large global bank to include but not limited to capacity planning, rule review, source to surveillance mapping and enhancement of coverages. Activities included current state analysis and optimization using an analytics environment in a secure data lake built on a big data architecture. He provided strategic recommendations around rule retention and applicability based on the bank's typologies.

As a senior architect, Sang has led the design, development, and implementation of a custom monitoring solution for a large global bank to mitigate gaps identified in the institutions existing monitoring controls. He determined the bank's overall residual risk exposure and implemented the transaction monitoring solution cover the risk.

Sang has worked as a program manager for a top tier global bank to help establish the going forward strategy and manage the execution activities to include collaboratively working with stakeholders to establish turnkey milestones and budgeting efforts for a 3 year program to overhaul the transaction monitoring solution across 30+ countries for all lines of businesses as part of a financial crime risk mitigation program.

He has worked with a top tier global bank in Canada and Mexico to perform an end to end testing and optimization of their transaction monitoring system that included data lineage analysis from source system to the monitoring tool, testing of the scenario logic and implementing segmentation and tuning methodology to optimize the alerts being generated by the tool.

He worked with several organizations to strategize future state design and architecture for client’s transaction monitoring system through business requirement definition discussions with critical stakeholders. He supervised the design and implementation of transaction monitoring system and case management tool for managing and reporting suspicious money laundering transactions for a bank holding company’s auto-financing business and the online banking business in the United States and in Canada.