Top Of House Reporting and RPA in Internal Audit using Big Data

Paresh Patel  CFE, CIA, CISA, CRISC, CRMA, FCCA
Audit Analytics Manager

6/5/2018
Learning Objectives

- Data Analytics & Auditing
  - Data Types
  - Traditional Data Analytics Audit Applications

- Tools and Technologies
  - Open Source Vs. Paid

- Modern Day Data Analytics Audit Applications
  - Robotics / RPA in Internal Audit
  - Artificial Intelligence in Internal Audit
  - Auditing RPA
WFC Company - 10 Business Divisions

Shared Services
- Chief Administrative Office (includes SCM, HR)
- Corporate Risk
- Law
- Enterprise Finance (includes Treasury) & Information Technology
- Audit Services

LOBs
- Consumer Lending
- Community Banking
- Payments, Virtual Solutions, and Innovation
- Wealth and Investment Management
- Wholesale Banking
Data Analytics and Auditing

1. Understand your business process, risks, controls
2. Gain access to the data
3. Build test cases/models (predict future real-world scenarios)
4. Perform analytics (models/algorithms) to help uncover patterns within the data
5. Uncover hidden patterns and relationships in the internal data, combine with external data
6. Act on the insights

Retroactive and intuitive >> proactive and data-driven

Balance the risks and opportunities of the analytics to be performed
Data – typical sources

Internal (generated by the entity)

- **Master Data**
  - Customer / Patient/ Vendor /Product

- **Transactional data**
  - Sensor / POS data
  - Patient Medical Records
  - Call Monitoring data
  - Sales/ Invoicing /Billing

- **Shared Services Data:**
  - Financial: GL, JE, etc.
  - HR: employees

- **Reference Data**
  - Hierarchy/ Rollup /Code /Translations/ Mapping/ AU, Chart of Accounts Hierarchy, etc.

- **Unstructured Data**
  - Emails, PDF/Word documents, XML

External (generate outside the entity)

- **Industry /Trade data**
  - Broker data
  - Industry specific sentiment data
  - NAICS
  - Trend data

- **Social Media data**
  - Facebook
  - Twitter

- **Census Bureau Data**
  - Race, Ethnicity, Income

- **Open Data**
  - Local county/ city /state /federal
  - Statistical data from Labor, Environmental, etc.

- **Research**
  - Healthcare Quality /Statistics

- **Subscriptions databases**
  - Pitney Bowes
  - Dun & Bradstreet
# Traditional Data Analysis Audit Applications

<table>
<thead>
<tr>
<th>Audit Technique</th>
<th>Benefits</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Integrity / Data Lineage</strong></td>
<td>Provides end-to-end assurance that data is accurate and complete</td>
<td>Manual process that is very time consuming</td>
</tr>
<tr>
<td><strong>Report Validation</strong></td>
<td>Ensures information reported to key stakeholders is substantiated</td>
<td>Manual process that is very time consuming</td>
</tr>
<tr>
<td><strong>Code Validation</strong></td>
<td>Ensures that if executed as planned, code will return proper results</td>
<td>Manual process that is very time consuming and open to manipulation of code or results</td>
</tr>
<tr>
<td><strong>Data Visualizations</strong></td>
<td>Allows stakeholders to view difference slices and presentations of data</td>
<td>Manual process that is very time consuming to create and update the visualizations</td>
</tr>
<tr>
<td><strong>Data Exploration</strong></td>
<td>Ability to find new insights to find results that impact audit planning and findings</td>
<td>Manual process that is time consuming and may or may not produce results</td>
</tr>
</tbody>
</table>
Tools and Technologies – what to choose?

It all depends...

Pick the tool /language that best fits the needs. Factors:

- What **problems** do you want to solve? Are you looking for data exploration, acquisition, analysis, reporting/visualizing, etc.?
- What is the **size** of data, variety of data (structured vs. unstructured), etc. you anticipate?
- What are the initial and ongoing **costs** for buying, learning, keeping up with the changes, licensing, etc.?
- What are the commonly used **tools** in your field /industry?
- What are the other tools are available in your **organization** that you can 'tap' into?
Tools and Technologies

Tools for data

- Analysis
- Acquisition
- Wrangling
- Visualizing
- Reporting
- Automation

Paid Vs. Open Source (free)
Tools and Technologies – Paid Vs. Open Source

Excel/Access (Paid-sort of)

- Part of MS Office on Windows and Mac machines
- Visually see the data
- Newer versions can handle more data (1MM Rows)
- Rich connectivity to the Microsoft stack of technologies e.g. SQL Server, SharePoint sites, etc.
- Tons of 'add-ons' e.g. PowerPivot, PowerBI, Excel Add-Ins (e.g. SAS Add-In), etc.
- **Scripting** using Macros /VBA
Tools and Technologies – Paid Vs. Open Source

R (created in 1995):

• It’s a **free** software programming language and software environment (IDE)
• For **statistical** computing and **graphics** techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others.
• **Ease** of use and **extensibility** has raised R’s popularity substantially in recent years
Tools and Technologies – Paid Vs. Open Source

Python (created in 1991):

• It’s a **free** full featured popular programming language

• **Easy to learn** and get started (well not quite)

• Many **libraries** /packages, guides, tutorials

• Have to write code (analysis is **reproducible**)

• Lots of ‘**free**’ code already out e.g. on GitHub, blogs, forums, etc.
Paid Alternatives:

**SAS**: PC SAS, SAS EG, SAS VA, SAS Ent. Miner

Others include: Matlab, Azure ML, IBM SPSS, etc.

**Pros:**
- Lots of **features** / bells and whistles
- Includes all **packages** required
- Large scientific **community** (used by many universities and large companies)

**Cons:**
- Algorithms are **proprietary** (trust that they were implemented correctly)
- **Expensive** (used by people with sufficient funds to buy a license)
- **Portability** (the ability to run your code on someone else’s computer)
- Few third party **extensions** (compared to open source)
Modern Day Data Analysis Audit Applications

- Traditional Data Analytics
- Continuous Auditing
- Robotic Process Automation*
- Artificial Intelligence*

* This presentation will explore these topics and their application to internal audit further.
WHAT IS ROBOTICS?

Robotics is software that can navigate across different IT systems to work the way a human does.

- Robotics process automation (RPA) is the most commonly adopted type of robotics solution thus far, but cognitive computing solutions provide promise for the future.
- RPA automates heavily rules based processes with structured data; Cognitive computing automates unstructured data.

Key Terms

Robot = “Bot”
- An individual agent or “virtual worker” capable of performing tasks assigned to it.
- Alternatively with some tools, a bot can refer to the automated task itself, and the individual agent as a “runtime license.”

Configuration
- The action of creating an automation for a process or activity within a robotics software platform.
- Alternatively, some companies may use the terms “program,” “develop,” or “train.”

Vendor/Tool Examples
- Automation Anywhere, Blue Prism, UiPath, Workfusion, IBM Watson, Amelia, NICE, Redwood, Pega, Kofax

Source: CEB analysis.
RPA definition

RPA is the programming /coding of basic tasks that humans do.

RPA is designed to take away the human component (activity, step, etc.) and reduce the burden of repetitive, simple tasks on employees when processes require no judgment.

In other words, the programming can be coded as an automated workflow with the ability to do several types of tasks in order, across multiple data sources or across different platforms/applications.

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduces manual interaction into processes</td>
<td>Can be costly to implement</td>
</tr>
<tr>
<td>Provides more accurate results</td>
<td>Additional software capabilities may be required</td>
</tr>
<tr>
<td>More efficient and effective execution</td>
<td>Additional skills and resources may be required</td>
</tr>
</tbody>
</table>
Benefits of Automation

- Increase Coverage
  - Avoid Sampling – provide full population testing
- Reduce overhead once set / Zero errors if coded correctly
- Increase Visibility
  - Put resourced towards other work
  - ‘real time’ / any frequency of refresh
- Proactive / Emerging Risk identification
RPA – Internal Audit Application

**Full Population Testing**
- Useful for testing high risk areas where a minimal number of exceptions could have significant impact to Company performance / reputation
- Fraud testing – AP, vendors, employee expenses
- P &P testing e.g. password testing, change management, patch management, configurations management
- User Access testing- new users and terminations against HR data
- Journal Entries

**End to End / Reporting Testing**
- Ensure complete and accurate information is transferred from systems of record to regulatory and risk reports
- Data validation and quality, Control point testing, data lineage rules
- KYC, Vendor setup, GL Reconciliation, other calculations

**OCR / Document Reading**
- Scrape customer and contract information from a scanned or imaged document into a searchable database to perform full population or end to end testing

**Other Misc. Opportunities**
- Data parsing: e.g. data parsing, slice and dice using VBA automation
- Handling MS Outlook mailbox- save attachments, emails, forward, etc. using VBAs
- Data downloads and uploads
- Create Metadata e.g. dynamic data dictionaries from DBs
An example of RPA
Artificial Intelligence (AI) or cognitive technologies in audit can be applied as:

“Managing risk(s) using a variety of data, tools, and technologies to help drive proactive identification of control breakdowns as well as efficiently identifying and effectively managing emerging risks within a given process.”

AI = concepts and ideas related to development of intelligent systems

AI applications within internal audit:

- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics
- Natural Language Processing (NLP)
- Text Analytics
- Machine Learning

Utilizing these data analytics techniques, internal audit can identify emerging risks to drive audit planning, scoping, and testing
AI application - Descriptive Analytics

Descriptive analytics (aka BI) show us the landscape of using data at face value and can be shown through dashboard reporting and visualizations.
Predictive Analytics or prescriptive analytics take past data and apply advanced analytic concepts to identify emerging risks and direct audit teams real time where to apply resources.

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Utilize the <strong>textual</strong> data consumers are saying about you on social media to identify areas that may indicate <strong>potential reputation</strong> risk to the company.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Alerts</td>
<td>Monitor <strong>key customer service</strong> data such as cases by products to indicate areas that may have weaker controls, lack of governance, or fraudulent activity.</td>
</tr>
<tr>
<td>Industry data</td>
<td>Harvest <strong>publically /OpenData</strong> available for your industry to identify areas where losses are occurring for peer competitors to identify areas that may impact your organization. Data may include regulatory enforcement actions, peer losses, and census data.</td>
</tr>
</tbody>
</table>

*Source: brand24*
AI applications - NLP

https://www.youtube.com/watch?feature=youtu.be&v=D5VN56jQMWM&app=desktop

How do we apply NLP in internal audit?

**Text:**
Read and understanding free form text in key documents including contracts, policies and procedures, confirmations, and social media sentiments

**Speech:**
Capturing attendance at key meetings

Listening and interpreting customer service calls
AI applications – Text Analytics

Text Analytics allow you to take the vast amount of textual data available and develop relevant insights, testing, analytics, and visualizations based on the data.

| **Audit Reports** | Review audit reports and identified issues for key themes that may exist across the organization to get a horizontal perspective of the risk management environment |
| **Reputation Risk** | Review social media posts, news articles, and other public documents for potential reputation risk concerns such as political, customer harm, and litigation |
| **Journal Entries** | Reviewing journal entries for keywords that may indicate fraud such as manual adjustments, overrides, and suspicious entry descriptions |
| **Document Reading (OCR)** | Review documents (vendor contracts, customer contract, invoices, loan documents, etc.) to ‘read’ and capture data for consumption (building dashboard, building analytics, etc.) |
AI applications - Machine Learning

First let’s define ML:

*Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.*

It focuses on solving real-world problems by mimicking our decision-making logic.

Given that, apply non-linear techniques to use cases using historical data to model and real-time data to learn e.g.

- **Monitoring and Fraud detection in Expenses** – let the AI drive thresholds, time periods, and employees to focus on.
- **Monitoring and Fraud detection in AP** – let the AI drive which Txns or approvers to test on – given volume changes, high-dollar Txns, etc.
- **Monitoring using advanced text analytics** – let the AI drive what keywords to ‘scope-in’ for analysis.
- **Monitoring in Underwriting** – let the AI look at loans issued recently (that used traditional parameters) that resulted in default.
AI applications - Machine Learning

Main objective is to build algorithms that can receive input data and use statistics to predict the outcomes.

Industry applications for various tasks:
• Recognizing objects*
• Translating speech in real-time*
• Determining potential outcomes
• Understanding consumer habits**
• Making personalized recommendations**
• Fraud detection, cybersecurity and tax evasion – ML algorithms analyze historical transactions data, social network information, and other data to recognize patterns and spot anomalies
• Predictive maintenance in IoT - gather data about everyday objects, such as fuel gauges and tires, and share it across the network, to predict performance and further outcomes
• Service personalization - combine historical customer service data, NLP, and ML algorithms that continuously learn from business interactions

*Apple, Google, and Microsoft
**Amazon, Netflix
Auditing RPA

- Strategy
- Governance
- Architecture and Infrastructure
- Human Factor
- IT controls
- Data Quality
- Ethics and Privacy
END_PROGRAM

@ Paresh.Patel@wellsfargo.com