Using Predictive Analytics to Detect Contract Fraud, Waste, and Abuse
Case Study from U.S. Postal Service OIG

MACPA Government & Non-Profit Conference
April 26, 2013

Isaiah Goodall, Director of Business Development
goodall@datamininglab.com
540-560-3183
A Brief Introduction To Elder Research, Inc.

• Largest, most experienced consultancy in Data Mining and Predictive Analytics (founded in 1995)

• Experts in applying COTS data mining products (SAS, SPSS, etc.) to solve real business problems

• CEO Dr. John Elder is an award-winning author on data mining, adjunct professor at UVA, and keynote speaker at large analytics conferences (KDD, Predictive Analytics World, etc.)

• Specialize in applying predictive analytics and text mining to fraud detection

• Over 16 years of data mining experience for 75+ clients, including 10 Federal Agencies
ERI Authored Books

Book written for practitioners by practitioners (May 2009)
2009 PROSE Award for Mathematics Winner

How to combine models for improved predictions (Feb 2010)

Published Jan 2012
What is Predictive Analytics?

- Predictive Analytics: Discovering patterns in past data that can be used to predict the outcome of future cases.
  - Uses anomaly detection and statistical algorithms to find patterns and anomalies in large amounts of data.
Needles in the Haystack

- Use predictive analytics to remove 90% of the hay to focus on the 10% with the most needles
- Augments analysts & investigators helping them focus on highest ROI leads
- Reduces false positives by scoring and ranking suspicious cases
The 9 Levels of Analytics

Descriptive Techniques:

1 – Standard Reporting
   “How much did we sell last quarter?”

2 – Custom Reporting or “Slicing and Dicing” the Data (Excel)
   “How many investigations did we perform in each state last year?”

3 – Queries/drilldowns (SQL, OLAP)
   “Which contractors received over $10 million in sole-source contracts last year?”

4 – Dashboards/alerts (Business Intelligence)
   “In what sectors have customer complaints grown since last quarter?”

5 – Statistical Analysis
   “Is frequency of communication with the customer correlated with satisfaction?”

6 – Clustering (Unsupervised Learning)
   “How many fundamentally different types of behaviors are in the data and what do they generally look like?”

Predictive Techniques:

7 – Predictive Modeling
   “Which contracts are most likely to be fraudulent?”

8 – Optimization & Simulation
   “What number of investigators would we put on each case to maximize expected return?”

9 – Next Generation Analytics – Text Mining & Link Analysis
   “Do the transactions reveal a coordinated set of people likely to be a fraud ring?”
Benefits of Using Predictive Analytics for Contract Fraud Investigations

- Examine random samples
- Use single data sources
- Reliance on tips, hotline
- Pay-and-chase
- Reliance on hunches, past experience only
- Known fraud schemes only
- Subjective consideration of different suspicious activity

- Examine and score each contract
- Unified view of contract (cradle to grave)
- Systematic way of generating leads
- Preventing fraud/stemming it early
- Data-driven decision making
- Identify emerging fraud schemes early
- Assign weights to fraud indicators mathematically
The Data Mining Process

Elder Research, Inc.
DATA MINING FOR FRAUD

Analyzing populations (the haystack) instead of individuals (leaves of hay). **Needles** should look different when compared to the population of normal leaves of hay ...
Hopefully needles look different. Should we mention that this is a possibility i.e. that they all look the same.
UNDERSTAND THE PROBLEM

• What data is available?
• What processes create the data?
• Who are the experts?
• What are the known schemes or cases (...what do the needles look like)?
WHAT IS THE POPULATION?

- Is the population large enough?
- Are we working with a sample?
- Does the population change over time?
- How many features (or columns)?
By identifying the schemes we are forcing the bad guys to change their approach which in turn changes the population.

THOMAS HEIMAN, 1/26/2012
SUPERVISED VS. UNSUPERVISED

- HC Claimant Model
- HC Provider Model

<table>
<thead>
<tr>
<th>Name</th>
<th>Bal</th>
<th>Fraud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob Jones</td>
<td>234</td>
<td>N</td>
</tr>
<tr>
<td>Bill Barney</td>
<td>87</td>
<td>Y</td>
</tr>
<tr>
<td>Sheri Detmier</td>
<td>25367</td>
<td>N</td>
</tr>
<tr>
<td>Ocho Cinco</td>
<td>3423</td>
<td>N</td>
</tr>
</tbody>
</table>

- SS&E Model
- Transportation Model
- Stamp Stock Model
- Mail Theft Model

<table>
<thead>
<tr>
<th>Name</th>
<th>Bal</th>
<th>Fraud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob Jones</td>
<td>234</td>
<td>?</td>
</tr>
<tr>
<td>Bill Barney</td>
<td>87</td>
<td>?</td>
</tr>
<tr>
<td>Sheri Detmier</td>
<td>25367</td>
<td>?</td>
</tr>
<tr>
<td>Ocho Cinco</td>
<td>3423</td>
<td>?</td>
</tr>
</tbody>
</table>
**ASSESSING A SUPERVISED MODEL**

Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F</strong></td>
<td>800,000</td>
<td>20,000</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>100</td>
<td>900</td>
</tr>
</tbody>
</table>

*workload*
UNSUPERVISED MODELS

• Set of rules
  • Was contractor terminated for default?
  • Is CO also COR?
  • Are invoices submitted at abnormal frequencies?
  • Has the contract gone through excessive modifications?
ASSESSING UNSUPERVISED

• Manually investigate cases with highest risk scores
• Determine case status: Fraud, Internal Control, Management, Normal
• Feedback loop: input outcome into back into model
  • Model “learns” from new outcomes
WRAP UP

• Analyze “normal” populations (what features are important)
• Look for anomalies and signals that describe the needles (iterate through multiple experiments)
• Use feedback and known cases to improve needle detection
• Knowing a case is NON-fraud is as important as knowing a case is fraud!

• QUESTIONS?
Case Study: USPS OIG Contract Fraud

- $73 billion in revenues
- $900 billion mailing industry
- 800,000 employees and contractors
- 37,000 postal facilities
- 213 billion pieces of mail delivered
- $33 billion in contracts (FY2009)

**Mission**: Maintaining the integrity and accountability of America’s postal service, its revenue and assets, and its employees.
Contracting with the USPS

USPS managed $33 Billion in contracts (FY2009)
Contract Fraud at USPS OIG

Elder Research was engaged by USPS OIG to apply advanced analytics to support the Contract Fraud Program

• What is Contract Fraud?
  – A deliberate deception in order to secure unfair or unlawful gain, that adversely affects the Postal Service’s interests

• Who commits it?
  – Supplier/contractor, contracting officer, contracting officer's representative, postal management
  – Working individually or in collusion
USPS OIG Contract Fraud

Problem Description

- In FY 2009 the Postal Service managed over $33 billion in postal contracts.
- Reliant on tips (reactive) - need a proactive way of providing good leads to investigative team.
- Wide variety of fraud schemes; moving target
  - Need a model that can be updated
  - Continuous Monitoring
Challenges in Detecting Contract Fraud

- Disparate data sources
- Data cleanliness
- **Few known cases** of fraud
- **Wide range** of fraud schemes
- Necessary data not always captured
- **Long investigative times** to validate leads
- Cost of private data (D&B, Lexis Nexis, etc.)
- **Buy-in** from investigators and auditors (culture change)
Custom Risk Indicators to Score Every Contract

- ERI Solution: Contract Fraud Detection Model
  - Generates leads based on risk indicators and anomaly detection.
  - 30 + custom fraud indicators or “red flags” developed
  - Contracts scored and rank ordered based on weighted combination of the indicators
Three Categories of Fraud Indicators

### Mods / Contract

| MC1 | High number of mods per contract category |
| MC2 | High number of mods (Bayesian adjusted for Supplier) |
| MC3 | Mod #s in sequence |
| MC4 | Time between 1st contract and mods (30 days) |
| MC5 | Mod % change from original contract (then aggregate by supplier/employee) |
| MC6 | Mod date before award date |
| MC7 | Contract amount is zero, but not cancelled |
| MC8 | indicator for sole-source/non-compete contract |
| MC9 | termination for default |
| MC10 | Outlier by cost for product category |

### Payments

| PY1  | Duplicate Invoices |
| PY2  | Frequency of payment to suppliers/by contract |
| PY3  | Payments to multiple suppliers at same address |
| PY5  | Payments to multiple suppliers for same contract |
| PY7  | Payments exceed contract amount |
| PY8  | Payment going to IRS or state agency |
| PY10 | ID source system = Manual invoice entry (% of entries that have that) |
| PY11 | Line Amount Outlier |

### Personnel

| PL1  | CO acting outside authority |
| PL2  | CO has authorization to do non-compete |
| PL3  | CO splitting costs to avoid higher level review |
| PL4  | CO-Supplier Relationship Strength |
| PL5  | # of times CO is COR, Buyer, Requestor |
| PL6  | Unusual use of annual leave |
Scoring System

- **Flag**
  - Was the metric tripped?

- **Percent**
  - How severely was the metric tripped?

- **Rank**
  - Based on dollar amount, how important is it?

Combine

Final Risk Score
USPS OIG Results From Initial Model

• Forensic examiners evaluated each metric for its usefulness and weighting in the overall model.

• **23 of 31** top leads (highest scored contracts) were useful:
  – Office of Investigations
  – Office of Audits
  – Postal Management

CF Model: 74% useful
Fundamental Shift in Approach

• **Proactive** rather than reactive
• **A single tool** that serves multiple stakeholders:
  – Investigations
  – Audit
  – Inspections
  – Management
  – Board of Governors
• **Buy-in from end users**
  – Get them on your team
  – Intuitive, easy-to-understand tool that gives valuable leads
High-Level Model Architecture

- Contract Data
- Payments Data
- Employee Data
- External Data (EPLS, D&B)

CF Model

RISK SCORES
<table>
<thead>
<tr>
<th>Contract ID</th>
<th>Risk Score</th>
<th>Amount</th>
<th>Mods</th>
<th>Current CO</th>
<th>Start/Sched. Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0DJFYW47W8060</td>
<td>86</td>
<td>$0.00</td>
<td>11</td>
<td>BAKER, JOSHUA</td>
<td>N/A - N/A</td>
</tr>
<tr>
<td>4NPDBB96F9700</td>
<td>86</td>
<td>$228,200.00</td>
<td>9</td>
<td>HUGHES, JOSHUA</td>
<td>11/01/2007 - 10/31/2011</td>
</tr>
<tr>
<td>0ZUZKH59B0749</td>
<td>71</td>
<td>$27,757,798.00</td>
<td>51</td>
<td>POTTS, MARJORIE</td>
<td>07/31/2007 - 09/30/2014</td>
</tr>
<tr>
<td>4ZMUAPA42K2667</td>
<td>70</td>
<td>$35,553,028.00</td>
<td>56</td>
<td>WHITE, JUANITA</td>
<td>11/26/2007 - 12/31/2011</td>
</tr>
<tr>
<td>0FPN5143F7921</td>
<td>68</td>
<td>$27,781,666.00</td>
<td>39</td>
<td>BRIGGS, JAMES</td>
<td>06/01/2010 - 07/31/2013</td>
</tr>
<tr>
<td>4RGHLQ33P5153</td>
<td>68</td>
<td>$200,000,000.00</td>
<td>9</td>
<td>COLLINS, DEBORAH</td>
<td>12/21/2002 - 12/19/2012</td>
</tr>
<tr>
<td>4YPM5397979</td>
<td>65</td>
<td>$68,000,000.00</td>
<td>17</td>
<td>SWANSON, JOLENE</td>
<td>04/02/2007 - 04/01/2014</td>
</tr>
</tbody>
</table>
## Contract 9HBM1U78M0027

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Amount</th>
<th>Mods</th>
<th>Current CO</th>
<th>Start/Sched. Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOLUTIONS &amp; EXCAVATING ELLIS</td>
<td>$60,495,844.00</td>
<td>32 WILLIAMS, LOREN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Status**: Open  
**Portfolio**: MAIL AND OPERATIONAL EQUIPMENT  
**Issuing Office**:  
**Administering Office**:  
**Purchase Description**: Specialized warehousing and storage  
**Award and Modifications**

<table>
<thead>
<tr>
<th>Num.</th>
<th>Action Reason</th>
<th>Amount</th>
<th>Comp. Class</th>
<th>System Create Date</th>
<th>CO</th>
<th>Start Date</th>
<th>Sched. Compl. Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awd</td>
<td>N/A</td>
<td>$19,899,538.00</td>
<td>A</td>
<td>01/27/2004</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>3</td>
<td>Administrative</td>
<td>$0.00</td>
<td>A</td>
<td>03/30/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>4</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>03/30/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>5</td>
<td>Administrative</td>
<td>$0.00</td>
<td>A</td>
<td>04/25/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>6</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>08/11/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>7</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>08/11/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>8</td>
<td>Administrative</td>
<td>$0.00</td>
<td>A</td>
<td>09/06/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>9</td>
<td>Administrative</td>
<td>$5,973,411.00</td>
<td>A</td>
<td>11/22/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>10</td>
<td>Administrative</td>
<td>$0.00</td>
<td>A</td>
<td>12/09/2003</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>11</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>01/20/2006</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>12</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>02/03/2006</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>13</td>
<td>Administrative</td>
<td>$0.00</td>
<td>A</td>
<td>05/18/2006</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>14</td>
<td>Supplemental/Bilateral</td>
<td>$0.00</td>
<td>A</td>
<td>07/19/2006</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
<tr>
<td>15</td>
<td>Supplemental/Bilateral</td>
<td>$5,920,704.00</td>
<td>A</td>
<td>10/26/2006</td>
<td>Reeves, John</td>
<td>03/17/2004</td>
<td>03/31/2007</td>
</tr>
</tbody>
</table>
## Contract 9HBMIU78M0027

**Supplier**: SOLUTIONS & EXCAVATING ELLIS  
**Amount**: $60,495,844.00  
**Mods**: 32  
**Current CO**: WILLIAMS, LOREN  
**Start/Sched. Completion**: 03/17/2004 - 09/31/2011

### Modification Timeline

<table>
<thead>
<tr>
<th>Amount</th>
<th>Create Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>$27,600,000.00</td>
<td>2004-01</td>
</tr>
<tr>
<td>$37,600,000.00</td>
<td>2004-02</td>
</tr>
<tr>
<td>$40,400,000.00</td>
<td>2004-03</td>
</tr>
<tr>
<td>$42,200,000.00</td>
<td>2004-04</td>
</tr>
<tr>
<td>$43,900,000.00</td>
<td>2004-05</td>
</tr>
<tr>
<td>$46,700,000.00</td>
<td>2004-06</td>
</tr>
<tr>
<td>$49,500,000.00</td>
<td>2005-01</td>
</tr>
<tr>
<td>$54,000,000.00</td>
<td>2005-02</td>
</tr>
<tr>
<td>$50,400,000.00</td>
<td>2005-03</td>
</tr>
<tr>
<td>$52,200,000.00</td>
<td>2005-04</td>
</tr>
<tr>
<td>$54,400,000.00</td>
<td>2005-05</td>
</tr>
<tr>
<td>$56,700,000.00</td>
<td>2006-01</td>
</tr>
<tr>
<td>$59,400,000.00</td>
<td>2006-02</td>
</tr>
<tr>
<td>$43,900,000.00</td>
<td>2007-01</td>
</tr>
<tr>
<td>$42,200,000.00</td>
<td>2007-02</td>
</tr>
<tr>
<td>$40,400,000.00</td>
<td>2007-03</td>
</tr>
<tr>
<td>$38,300,000.00</td>
<td>2007-04</td>
</tr>
<tr>
<td>$36,200,000.00</td>
<td>2007-05</td>
</tr>
<tr>
<td>$34,300,000.00</td>
<td>2007-06</td>
</tr>
<tr>
<td>$32,400,000.00</td>
<td>2007-07</td>
</tr>
<tr>
<td>$30,500,000.00</td>
<td>2008-01</td>
</tr>
<tr>
<td>$28,800,000.00</td>
<td>2008-02</td>
</tr>
<tr>
<td>$27,100,000.00</td>
<td>2008-03</td>
</tr>
<tr>
<td>$25,600,000.00</td>
<td>2008-04</td>
</tr>
<tr>
<td>$24,200,000.00</td>
<td>2008-05</td>
</tr>
<tr>
<td>$22,900,000.00</td>
<td>2008-06</td>
</tr>
<tr>
<td>$21,600,000.00</td>
<td>2009-01</td>
</tr>
</tbody>
</table>

## Contract 4TSFGP54U2233

**Supplier**: COLUMBIA WELDING SMITH TRANSPORTATION AND  
**Amount**: $0.00  
**Mods**: 13  
**Current CO**: BAKER, JOSHUA  
**Start/Sched. Completion**: N/A - N/A

### Contracts with similar risk

<table>
<thead>
<tr>
<th>Contract</th>
<th>Risk Score</th>
<th>Dollar</th>
<th>Contract Size</th>
<th>Same Invoice Diff Codes</th>
<th>Line Items Not Equal Invoice Unit</th>
<th>MCO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6B</td>
<td>64.4</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>8.0</td>
<td></td>
</tr>
</tbody>
</table>

### Supplier Summary

- **SOLUTIONS & EXCAVATING ELLIS**
  - TIN: 197797426
  - 46782 North Road, Suite 9719
  - Jeann, NV 89026

- **Williams, Loren**
  - Office in Huntsville, AL 35806
  - Authorized for $106,000.00
  - Non-compete: $350,000
  - 35 Contracts
  - Riskiest Contracts
    - 6B | 68.4 | $70.0 | 6B | 68.4 | $70.0 |

### CO Summary

- **Williams, Loren**
  - Office in Huntsville, AL 35806
  - Authorized for $106,000.00
  - Non-compete: $350,000
  - 35 Contracts
  - Riskiest Contracts
    - 6B | 68.4 | $70.0 | 6B | 68.4 | $70.0 |
RADR - Return on Investment

• RADR deployed to over 1,100 investigators and auditors at USPS OIG

• Contract Fraud – $300,000 in recoveries in model validation phase

• Healthcare Fraud RADR (Workers’ Comp Fraud)
  – 113 cases initiated; 12+ cases resolved
  – Over $11 million in recoveries, restitutions, and cost avoidance in first year
  – Work hours per case – (30% decrease)
  – $ returned per case – (35% increase)
Take-Away Lessons

• Don't wait for “perfect” data
• Don't wait for the “perfect” technology solution and support
• Start with what you have
• Get a “quick win” with a small pilot project to prove value
• Build team of domain experts & data scientists
• Understand the power of visualization
Questions?

Contact Info:
Isaiah Goodall, Director of Business Development
Elder Research, Inc.
goodall@datamininglab.com
540-560-3183