Predictive Analytics for Government & Finance

Presenters:

- Moderator: Larry Sapp, Finance Manager, Hilton Head Public Service District #1
- Speakers:
  - Eero Kilkson, Chief Data Architect, City of Minneapolis
  - John Senegal, Architect/Chief Programmer, IBM

Date: May 19, 2014
Agenda

• Why do I care about Big Data and Analytics?

• How is it relevant to Government Finance?

• How are cities and industries using it?

• The Minneapolis Case study
Why do I care about Big Data and Analytics?
Big Data?

Extracting insight from an immense volume, variety and velocity of data, in context, beyond what was previously possible.

- **Variety**: Data is now available in many different forms and places, i.e., social media, city agency systems, federal systems, documents, blogs, etc.
- **Velocity**: Data streams in from physical sensors or people sensors or data resides in government systems or in documents.
- **Volume**: The amount of data can be paralyzing for human insight alone.
We all face Information Overload... But are Lacking Insight

- 12+ TBs of tweet data every day
- 30 billion RFID tags today (1.3B in 2005)
- 4.6 billion camera phones worldwide
- 25+ TBs of log data every day
- 100s of millions of GPS enabled devices sold annually
- 76 million smart meters in 2009... 200M by 2014
- 12+ TBs of tweet data every day
- 200M people on the Web by end 2011
So what is The Big Data Conundrum?

- The percentage of available data an enterprise can analyze is decreasing proportionately to the available to it
  - Quite simply, this means as enterprises, we are getting “more naive” about our business over time
  - We don’t know what we could already know….
Information is at the center of a new wave of opportunity

44x

as much Data and Content Over Coming Decade

2009
800,000 petabytes

2020
35 zettabytes

1 in 3
Business leaders frequently make decisions based on information they don’t trust, or don’t have

1 in 2
Business leaders say they don’t have access to the information they need to do their jobs

83%
of CIOs cited “Business intelligence and analytics” as part of their visionary plans to enhance competitiveness

80%
Of world’s data is unstructured

60%
of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions
Machines vs. Management

Training
Scoring

Machines  Management

Score = 0.65

Gut Feeling!
Data to Insights to VALUE

- Data is critical and should drive decisions.
- Tools to process data are useful but could be using outdated data that is only useful for reporting what happened.
- To become really valuable, data must be transformed into useful relevant insight.
Getting to Value

What Happened?

Why?

What Next?

Business Intelligence 80%

Predictive Analytics 250%

Quantifiable ROI
Predictive Analytics

Predictive Analytics = Advanced Analytics + Decision Optimization

- Statistics
- Data Mining
- Text Mining
- Visualization
- Reporting
- Scoring Engine
- Rules Engine
- Recommendation Engine
- Optimization Engine
Getting to Value

• 48% of firms using predictive analytics are able to gather and analyze unstructured data compared to 14% of companies not using predictive analytics.

• Firms using predictive analytics saw an 11% increase in the total number of customers compared to an 8% increase in firms not using this technology.

• Organizations using predictive analytics saw an 8% increase in cross-sell/upsell revenue compared to 3% for those companies not using the tools.

81% of companies were using analytics for customer profiling and profitability analysis compared to 62% of those with no predictive analytics.

63% of firms were using analytics for risk management compared to 52% of those with no predictive analytics.

63% of organizations were using analytics for regulatory compliance compared to 33% of those companies not using predictive analytics.

63% of businesses were using analytics for fraud detection and management compared to 29% of those companies not using predictive analytics.
Why should you care?

Why government needs to do more with less

by Tammy Kulesa

DO MORE WITH LESS #2014

LESS IS MORE #crop130
There’s Help For that!

- Predictive Analytics starts the process of helping with the “more for less” epidemic

- Allow the machine to wade through historical data to find and describe:
  - Patterns
  - Relationships
  - Anomalies

- Because of the variety, velocity and volume, it is almost impossible for humans to detect

- Allow the machine to wade through historical data to predict and prescribe
Where do we come from, where do we go?

Traditional Approach
- Lack of Insight
- Velocity
- Inefficient Access
- Volume
- Variety
- Inability to Predict
- Sense and respond
- Instinct and intuition
- Skilled analytics experts
- Back office
- Automated

New Approach
- Predict and act
- Real-time, fact-driven
- Everyone
- Point of impact
- Optimized
OKAY! I get it computers can do things I can’t.

What does that really mean to me as a Government finance person?
I have an idea!

How about you just increase the city’s monthly revenue without raising taxes?
I have an idea!

Using Advanced Analytics a city of a half a million people in the Southwest was able to increase their monthly revenue somewhere between an estimated $25 – $50K.
They did it by detecting Unlicensed Businesses

- Many cities estimate that between 5% and 15% of the businesses operating within their boundaries fail to properly register for all necessary permits and taxes, costing cities millions of dollars in lost revenue.

- Many existing business license management operations and analysis system have the following challenges:
  
  Spreadsheet-based analysis with a large number of ad-hoc procedures based on individual capabilities and knowledge.
  
  Databases that are highly proliferated with increasing volumes of data across a large number of applications.
  
  Slow response time to fraudulent events with increasing liability to these events.
  
  Lack of collaborative capabilities with other areas of the city.
Finding Missed Revenue for the City

Problem
Many cities spend a significant amount of time auditing businesses to confirm they have the correct licenses and permits. However, data is stored in multiple places and departments, limiting the ability of auditors to properly investigate unlicensed businesses.

Delivered Value
Cities improve current investigation practice by including all related city permit, tax, and records data and fraud patterns are quickly uncovered.

Shared intelligence and collaborative working ensures that investigations can be completed more effectively, and with cumulative intelligence.

Improve productivity and auditability of the investigations team enabling more cases to be handled.

The solution adds intelligence & puts the relevant information into the right hands at the right time.
Fused data sources provide a more complete business profile

- **Company Name**
- **+ Company Owner**
- **+ Type of Company**
- **+ Current Location**
- **+ Business Status**
- **+ License Status**

**State Data**

**County Assessors**

**Utilities / Web**

**Internal Sources**

**Business Licenses**

**Complete Profile of Un-Licensed Business**

- **Fused data allows a complete business profile**
  - Initial business entities are fused with external and internal data sources to present a complete business profile for further analysis.
New insights

City in United States identifies numerous businesses operating without permits, effectively increasing potential sales tax and permit revenues by $32K per month.
Hmm... Interesting!

How else can Predictive Analytics help me?
## US Government Accountability Office (GAO) forum

<table>
<thead>
<tr>
<th>Program</th>
<th>Agency</th>
<th>Reported Improper Payment Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare Fee-for-Service</td>
<td>HHS</td>
<td>$29.6 billion, 8.5%</td>
</tr>
<tr>
<td>Medicaid</td>
<td>HHS</td>
<td>$19.2 billion, 7.1%</td>
</tr>
<tr>
<td>Medicare Advantage (Part C)</td>
<td>HHS</td>
<td>$13.1 billion, 11.4%</td>
</tr>
<tr>
<td>Medicare Prescription Drug</td>
<td>HHS</td>
<td>$1.6 billion, 3.1%</td>
</tr>
<tr>
<td>School Lunch</td>
<td>USDA</td>
<td>$1.6 billion, 15.5%</td>
</tr>
<tr>
<td>Earned Income Tax Credit</td>
<td>Treasury</td>
<td>$12.6 billion, 22.7%</td>
</tr>
<tr>
<td>Unemployment Insurance</td>
<td>DOL</td>
<td>$10.3 billion, 11.4%</td>
</tr>
<tr>
<td>Supplemental Security Income</td>
<td>SSA</td>
<td>$4.7 billion, 9.2%</td>
</tr>
<tr>
<td>Old Age, Survivors &amp; Disability Insurance</td>
<td>SSA</td>
<td>$3.2 billion, 0.4%</td>
</tr>
<tr>
<td>Supplemental Nutrition Assistance Program</td>
<td>USDA</td>
<td>$2.7 billion, 3.8%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>$98.6 billion</strong></td>
</tr>
</tbody>
</table>

Source: Estimates reported by OMB for fiscal year 2012.
“Data analytics can be used to:

• Identify patterns or trends, and identify when aberrations occur
• Data-mining and data-matching techniques can identify fraud or improper payments that have already been made
• Predictive analytic technologies can identify fraud and errors before payments are made, helping avoid “pay and chase” payment recovery
• Determine whether problems are widespread and systemic in nature
• Evaluate program performance and outcomes
• For example, accessing data to verify program eligibility, thus improving program integrity and payment accuracy”
Smarter Cities Capital Planning Analytical tools help Avoid Tearing up Streets Unnecessarily

Especially, after just reconstructing the road above it 6 months earlier!
Intelligent City Planning and Operations – Business Value

Improve City Budgets without Raising the Tax Burden on the Citizens

**Infrastructure Planning Optimization**
*Reduce the cost of infrastructure project planning and execution.*

- **Understand the state of the infrastructure to project cost in coming years:** Present planners with a view of the health of all the city infrastructure to allow cities to
- **Leverage analytics to Reduce Planning Cost:** Provide infrastructure planners with the tools to make correct decision about infrastructure project more quickly. Gain an 8 fold Reduction in planning cycle saving $150K+ annually in personnel costs
- **Align Projects Across departments to combine and reduce costs:** Coordinate plans across city agencies to reduce execution costs and disruption to citizens. Reduce the number of street closures by half
- **Aligning Infrastructure age so the ground is only dug up once to fix multiple systems:** Determine the optimal remediation actions to align infrastructure age so all infrastructure can be rebuilt at the same time. Reduce projects that require digging up a street by half with potentially savings of millions of dollars
- **Optimize remedies by budget, goals and funding sources to gain the most benefit from the available funds:** Run multiple what if scenarios with different optimization goals to determine the best way to spend infrastructure dollars.
Infrastructure Planning

Problem
Cities are very concerned that because of the budget shortfalls they are delaying the replace, rehab and repair decisions of key infrastructure assets that will eventually result in an increase in the failure of assets, in average age of assets, and .

Delivered Value
An innovative approach to cross agency planning by bringing together the concept of total lifecycle management of city infrastructure using descriptive, predictive and prescriptive analytics.
- Asset Performance Analysis – predictive & quantitative analysis of current performance of city infrastructure;
- Identify short (1-5 years) and long (10 – 100 year) term investment candidates;
- Perform sustainability analysis
- Investment Planning - Perform comprehensive planning for optimal operation and capital management

Solution Description
- Annual programs (eg snow cleaning, leaf pickup)
- Reactive maintenance (road repair, water main breaks, street light problems)
- Predictive / proactive maintenance (eg: inspections, hydrant flush)
- Asset replacement based on age, condition or failure
- Federal funding, regional expansion, new development, private-public partnership
- Improving quality of service
- Constant average asset life
- Asset lifecycle costing and analysis
- Optimal allocation of O&M vs Capital money
- Integrated asset management
Intelligent Infrastructure Planning takes an innovative approach to cross agency planning by bringing together the concept of total lifecycle management of city infrastructure using descriptive, predictive and prescriptive analytics. The solution leverages three pillars:

**Asset Performance Analysis** – Cross-agency predictive & quantitative analysis of current performance of city infrastructure; Scoring framework to enable identifying low performing assets

**Strategic Needs Assessment** – Identify short (1-5 years) and long (10 – 100 year) term investment candidates; Perform sustainability analysis

**Investment Planning** - Perform comprehensive planning for optimal operation and capital management
Overview of City of Cambridge, Ontario, Canada

- Population: 130,000
- Properties: 36,000
- Infrastructure: $1,600,000,000
- TCA Assets: 30,000
- Asset Components: 225,000
- Utilities: Roads, sidewalks, water, sewer ...
- Upper Tier Government: Water Supply & Sewage Treatment

“I would like to synchronize the end of life for road, water, storm, sewer so that when we go into a street, we put in a new infrastructure from ground up. This will help bring down the cross agency costs substantially and improve quality of service...”

Mike Hausser  
Director for Asset Management  
City of Cambridge

“Capital planning process in Cambridge takes 6 months and 12 people and is very exhausting. We need a system that can help improve the quality and speed of planning process...”

Kealy Dedman  
Director of Engineering  
City of Cambridge
On March 16th, City of Cambridge made a press release about the impact and value of PALM Solution

Staying on the cutting edge of technological advancements is key to the success of running a municipality. Private sector firms have recognized the value of research and development for many years, and we are taking it into the public sector in ways that are brand new – Doug Craig, Mayor, City of Cambridge

The IBM ‘Analytics for City Services and Safety’ (ACCESS) program now enables the City of Cambridge to compile, synchronize and better understand information from a variety of sources optimize capital spending to achieve the highest value in renewing our infrastructure. The result is better utilization of limited resources to sustainably improve quality of service – City of Cambridge
Benefits Fraud investigative challenges

- The scope of the benefits fraud challenge is huge:
  - Example: 3% of the $2.26 trillion annual U.S. health care expenditure ($68B) is lost to fraud
  - Example: Food stamp / card skimming (2% of benefit) fraud still amounts to at least $750M a year
  - Abuse is growing at a rate of over 5% per year
- Tremendous volumes of disparate data from multiple sources
- Fraud hidden in complex layers of transactional data
- Sophisticated methods used by fraudsters
- A variety of mediums to connect to: manual/human intelligence, paper trails, Internet, telecom and other electronic networks
- All data is not always searchable and accessible
- New technology can be expensive, difficult to install and learn

1 NHCAA, 2009
2 U.S. Department of Agriculture, 2009
IBM i2 solutions are designed to disrupt and defeat Benefits Fraud

- Unemployment fraud
- Healthcare fraud
- Food benefit card fraud and other prepaid card fraud
- Housing benefit fraud
- Disability fraud
- Undisclosed income or capital (e.g. state benefits, pensions, tax credits, savings and property)
Benefit-specific analysis capabilities

Date of service analysis

Fraudulent billings

Provider-beneficiary commonalities
Fraud Intelligence Analysis for pension fraud

- Leverage data across multiple networks and data sources including HR data, data from pension providers, public records data, email and social network data to perform rapid and accurate analysis of individuals and groups involved in pension fraud activity
- Determine if pension fraud is an individual or colluded groups
- Use Social Network Analysis and other IBM i2 capabilities to get a clear picture of medical events relating to disability and/or pension claims
- Analyze information from online networks such as Facebook to discover activities that can validate or disprove pension fraud claims
- Analyze fraud committed across the entire universe of pension plans:
  - Defined contribution plans
  - Defined benefit plans
  - Fraud committed by individuals receiving benefits
  - Fraud committed by employers
  - Fraud committed by insurance plan promoters
  - Fraud involving corrupt lawyers and doctors that aid in pension fraud
  - Under-funding through nondisclosure and misrepresentation
  - Oversight and compliance investigations (Employee Benefits Security Administration, Pension Benefit Guaranty Corporation, etc.)
City workers have described a variety of contexts in which to use analytics:

**Operational Planning**
- Predict the optimal allocation of resources
- Reduce duplicate effort among agencies

**Operations in Action**
- Share information in a timely fashion as workers “crawl the city”
- Know where to focus and shift resources right now

**Operational KPIs & Reporting**
- Measure and improve productivity & strategy
- Monitor overall health and vitality of an area

**Strategic & Long Term Planning**
- Resource capacity planning
- Infrastructure planning

**Event Planning & Permitting**
- Coordinate around an event or permit issuance
- Assess impact of doing or not doing something

**Exploring Data for Insights**
- Identify where & when to intervene
- Find non-obvious patterns in data
Together we are building Smarter Cities

**Leaders want:**
Data turned into actionable information

**IBM delivers:**
Better insight and decision-making

**Leaders want:**
Solution roadmap that aligns to my needs

**IBM delivers:**
A comprehensive and proven portfolio

**Leaders want:**
Real world examples to leverage & extend

**IBM delivers:**
Global experience with city engagements

Empowered leaders. Proven solutions. Real outcomes.
QUESTIONS?
LET'S TAKE A
DEEPER DIVE INTO
A SPECIAL CITY

YOUR HOST CITY
MINNEAPOLIS
IBM Brings it All Together

Data Driven Decisions

Open Data

Video

Sensors

Citizens

Smart Cities Platform

Silo’d City Departments

Social Media Analytics

Correlation

Prediction
Be SMARTER

Get data from:

- water sensors,
- open data sites,
- fire alarms,
- city vehicle GPS,
- train wheels,
- cameras,
- citizens, etc.
“Analyze statistics to see the trends, report on different performance metrics and find bottlenecks...”
Automate, Notify and Coordinate across agencies to make their work more affective...

To make sure everybody KNOWS WHAT TO DO!
Cities Fail when they do not Understand the Real Time Changes occurring in their Environment and they do not Respond to those Changes
Open Data Changes the Game ... Across the Globe
Minneapolis Case Study

GFOA Annual Conference
May 19, 2014
✓ City leaders seek intervention opportunities in their event-driven cities; thereby, requiring info from diverse systems and departments.

✓ Proprietary government data has the potential to reshape governance dynamics.

✓ Give management high integrity information, if possible knowledge, to run the place.

✓ Technology can enable a city to tell meaningful stories through data—where participants generate the content.

✓ Big data is not the key, analytic models and simulations drive the City outcomes we seek.

✓ Strategic use of information is the key to differentiating our city from any other—which will include new forms of information.

✓ A big key for city government to achieve its goals revolves around data and systems that are trusted, transparent and real-time.

✓ I usually get asked for killer apps, today I got asked for killer outcomes.
Project Focused On Some of the Key Elements

- Increase Collaboration
- Consolidated Data
- Consolidated Analytics
- Enable the Users To Generate the Analytics
- Create Analytics that are Usable By A Range of Users
Better coordinate city operations to gain efficiencies  
Deal more effectively with special events  
Improve handling of emergencies
"Working" Functional Concept

- Assess the Situation
  - Pattern Identification
  - Extract the “signal from the noise”
  - Monitor (Have conditions changed?)
  - Identify benchmarks to improve the identification of change (e.g. seasonality removed)

- Plan Your Actions
  - Capacity analysis
  - Optimize our resources
  - Predict the impact of our planning
  - BE PROACTIVE

- Situational Awareness
- Planning Solutions

- Knowledge Capture
- Efficiency Coefficient

- Capture Experience
- Measure Effectiveness

- Institutional Knowledge capturing
- Learning & classification

- How effective was our plan
- How effective were our processes
- Statistical analysis and reporting
- Trend analysis
Residents / visitors
Elected Officials
Department leaders and employees

Business view – Enterprise versus specific need(s)
Boundary focus – City-wide versus specific geography (ward, precinct, neighborhood, housing district, etc.)
Data visualization – map/graph versus time

Emphasize the value in having tools that can be re-used for different studies. Do not require new coding for each model. Must be agile
Challenge: Turning data into decisions

- Philosophy
  - Data → Information → Knowledge → Decisions → Outcomes

- Presently the City is largely focused on
  - Rear-view mirror perspectives
  - Macro-geography with some exceptions
  - One dimensional (based on data from one department)
  - Reactive response

- Current City data-driven efforts
  - Police CODEFOR
  - Results Minneapolis (1,250+ metrics)
  - Intelligent Operations Platform (IOP)
What we get today
Issues with Current Analytics

- Labor Intensive
- Static
- Old Data
Future Approach

- **How can we make it happen?**
  - Event correlation

- **What will happen?**
  - Route analysis
  - Weighted hotspot

- **What happened and why?**
  - Hotspot
  - Anomaly detection
  - GPS analysis
  - Pattern discovery

- **How are things going?**
  - Dashboard
  - Scheduled report

Moving up the analytics continuum
Analytics is the Key

Hotspot Detection
Workers search for places of more than usual interest, activity, or popularity

Route Tracking
Workers track vehicle routes

Event Correlation
In the sea of data points, what events occurred leading up to an outcome?

Pattern Discovery
We know that when X and Y happen Z is an outcome, where might it happen next?

Anomaly Detection
When does a trend depart from the normal pattern. Seasonality plays a significant role in Minneapolis
Benefits for Business

- Time savings (production, analysis, decisions, etc.)
- Thinking in 2,3,4,... dimensions
- Thinking across the enterprise and beyond the boundaries of the enterprise
- Move up the analytics continuum
- Better knowledge leading to better decision-making leading to better outcomes
- Acting proactively
- Objective analysis of the effectiveness of programs and policies
Use Case #1: Bad Landlords

**Hypothesis:** most housing issues are caused by a handful of landlords

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Who is a bad landlord?</td>
<td>☐ Discover characteristics of a bad landlord through event correlation and anomalies; compare against all landlords through pattern matching</td>
</tr>
</tbody>
</table>
Use Case #2: Vacant Properties

**Hypothesis:** There is a tipping point where a concentration of vacant properties begins affecting the economic development of an area.

**Challenge**
- Which neighborhoods have vacant property density affecting economic development?

**Technique**
- Determine hotspots of vacant properties; compare surrounding area economics (assessed property values, business income, etc.) to like areas of city.
Use Case #3: Effect of City Planning on Growth and Property Investment

**Hypothesis:** City programs such as “Great Streets” has a positive impact on neighborhood vitality

**Challenge**
- Identify the effectiveness of City Planning techniques
- Separate changes to overall economic upswing from the direct effects of a City action

**Technique**
- Identify comparable areas (control areas)
- Track the economic and neighborhood vitality indicators in the identified areas.
- Compare the test areas vs the control areas
Potential Disinvestment Symptoms

- Decrease in owner occupied properties
- Decrease in property values
- Increase in incidences of crime
- Decrease of businesses in the area of decline
- Vacant properties
- Unpaid utility bills
- Nuisance violations (uncut grass, garbage in yard or alley, ...)
- Frequent calls for service (Police, Fire, or EMS)
- Decrease in building permits
Potential Financial Consequences

- Decrease in tax base
- Decreased success rate of businesses
- Change in business types
- Decrease in employment opportunities
- City resource drain
- Property quality decline limiting future sales
- Decreased utility revenues not due to a decrease in utility consumption
# Use Case #4: Disinvestment

**Hypothesis:** By tracking a set of diverse criteria areas of disinvestment can be identified in time to be mitigated

## Challenge

- Do we recognize areas where there is significant disinvestment too late?
- Can we identify the factors that lend to disinvestment?
- Identify how and where the City should invest resources to address trends of disinvestment?

## Technique

- Establish trends for datasets from sources across City departments
- Continuously, and automatically, monitor departures from these trends
- Alert City department when early sign of disinvestment occur
Hypothesis: there is a fine line between teardowns generating new construction versus teardowns resulting in vacant lots

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>How to understand</td>
<td>Track teardowns and rebuilds geographically</td>
</tr>
<tr>
<td>disinvestment at a</td>
<td>over time</td>
</tr>
<tr>
<td>glance?</td>
<td>Correlate teardowns to outcomes</td>
</tr>
</tbody>
</table>
The Act of Analysis

- 10% workforce doing analysis more-or-less full time
- Analysis moving to the streets—at the point-of-work
- One day may be put in the hands of residents / visitors

- What’s analyzed varies over time
- Analytic models need to run real time—“what if” scenario
- Data elements will grow and contract over time

- Emphasizes value in having a product with generic, and thus, wide-spread application
Future Outcomes

- Real-time usage at time of decision
  - During *Results Minneapolis* meetings
- True *mission control* operation

- Public facing side
- Available to all city workers

- City government becomes “*compelling*” to residents
  - Evoking interest, attention, or admiration in a powerfully irresistible way
Intelligent Operations Platform (IOP) – Improving City Operations

IOP

Advanced Analytics
Anomaly Detection
Hotspot Detection
Event Planning
Alerting

Information Exchange

City Systems of Record
- Public Works
- Police Incidents
- Traffic Accidents
- Reg Svcs Permits
- Fire Incidents
- Non-City Agencies
- Citizens 311/911
- DID Events

Dashboards, Reports, Workflows with Secure Access
Thank You & Questions