Maximizing Return and Minimizing Cost with the Decision Management Systems
Critical customer and operational agendas continue to drive the use of predictive analytics

**Customer Analytics Agendas**
- Attract the ideal customer
- Grow share of wallet
- Service & Satisfy Customers

**Operational Cost Take-out Agendas**
- Drive operational performance
- Manage day to day operations
- Plan for operational success
Assess vulnerability – Predictive modeling with SPSS
Automating & optimizing decisions requires complementary technology
Maximizing Return and Minimizing Cost with the Decision Management Systems

CASE STUDY:
INSURANCE FRAUD
Combining predictive analytics with other techniques ensures optimal outcomes for complex use cases

- South Africa’s largest short-term insurance company
- More than 650,000 policy holders
- Assets under management of 17 billion South African Rand (US $2.4 billion)
- Market share greater than 22 %

**Instrumented**
- When a claim is submitted Santam captures data related to a number of key risk indicators

**Interconnected**
- The analytical engine uses a combination of business rules and sophisticated predictive models to assess claims for potential fraud and transfer them to the appropriate processing channel.

**Intelligent**
- By segmenting claims according to risk factors Santam can focus on investigating high-risk claims and catching fraudsters while rewarding good customers with fast settlement and better service.
Combining predictive analytics with other techniques ensures optimal outcomes for complex use cases (continued)

- Significant results in terms of fraud detection, customer service and return on investment

- Before
  - Minimum time to settle a claim was three days

- After
  - Low-risk claims can be settled within an hour
  - Customers with legitimate claims get much faster service.
  - Significantly reduced the number of claims that need to be assessed by mobile operatives, which will lead to considerable operational cost savings.

- Enhanced fraud detection
  - In the first month able to identify patterns that enabled us to foil a major motor insurance fraud syndicate
  - Within the first four months saved R17 million on fraudulent claims and R32 million in total repudiations

- Solution delivered a full return on investment almost instantly
What the business user sees is very different from what the data miner and decision system see to detect insurance fraud.

Front-line Rep only sees “Refer” at the point of impact.
Five components of analytical decision making

1. Problem Definition
2. Data
3. Rules
4. Predictive Models
5. Scenario Testing, Optimization, & Simulation
### Data Source Fields

<table>
<thead>
<tr>
<th>Data source</th>
<th>Field name</th>
<th>Measurement</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;PROCEDURE_RISK_GROUP&quot;</td>
<td>Nominal</td>
<td>HIGH, LOW, MEDIUM, SUPER_HIGH</td>
<td></td>
</tr>
<tr>
<td>&quot;QUANTITY_INDEX&quot;</td>
<td>Continuous</td>
<td>[0, 240]</td>
<td></td>
</tr>
<tr>
<td>&quot;SERVICE_TYPE&quot;</td>
<td>Nominal</td>
<td>Anesthesiology, Pharmacy, Radiology</td>
<td></td>
</tr>
<tr>
<td>&quot;SUBMIT_CHG&quot;</td>
<td>Continuous</td>
<td>[0, 500]</td>
<td></td>
</tr>
<tr>
<td>&quot;SUBMITTED_CHG_INDEX&quot;</td>
<td>Continuous</td>
<td>[30, 249]</td>
<td></td>
</tr>
</tbody>
</table>
### Manage Global Selections

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Include/Exclude</th>
<th>Remove</th>
<th>Lock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude OIG List</td>
<td>Exclude</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Health Care Claim Properties

#### Choose Who This Claim Area Applies to

#### Use Rules to Decide Which Action is Triggered

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Risk points</th>
<th>Sort</th>
<th>Remove</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 High ratio of submitted charges amount vs. average for sar</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Low ratio of net eligible amount / submitted charge amount</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Super high fraud rate for previous year for the procedure</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 High fraud rate for previous year for the procedure</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Remainder</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Use a Model to Decide Which Action is Triggered

<table>
<thead>
<tr>
<th>Sum of Points</th>
<th>Allocate to</th>
<th>Remove</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>Suspend</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>Hold</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Approve</td>
<td></td>
</tr>
</tbody>
</table>
## Health Care Claim Properties

### Choose Who This Claim Area Applies to

### Use Rules to Decide Which Action is Triggered

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Risk points</th>
<th>Sort</th>
<th>Remove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider is under investigation</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High ratio of procedure quantity per visit vs. average for spa</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomaly is detected for the claim</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Edit Rule

- **Description:** Anomaly is detected for the claim
- **Risk points:** 100
- **Expressions:**
  - Fraud_AnomalyDetect_Deploy = T

**Choose value...**
The image shows a graphical representation of a statistical model used to predict the risk of fraud in claims. The model is based on various predictor variables, as shown in the tree structure on the left. The predictor importance chart on the right indicates the relative significance of each predictor in determining the target variable IS_FRAUD. The most important predictors are those with the longest bars, such as 'FIRST_YR_CLAIM' and 'ELECTRONICCLAIM'. The least important predictors are those with the shortest bars.
Cluster Sizes

- Normal Service: 22.5%
- High Fraud Rate: 18.1%
- High Charge for a P: 27.1%
- High Total Submits: 22.8%
- High Out of Pocket: 9.5%

Size of Smallest Cluster: 234 (9.5%)
Size of Largest Cluster: 814 (27.1%)
Ratio of Sizes: Largest Cluster to Smallest Cluster = 2.87
Detect and stop fraud - SPSS

Suspended Claim Properties

Choose Who This Risk Factor Applies to

Allocate Using Segment Rules

- Allocate using rules
- Multiple Allocation
- Allocate randomly

Allocate to: First valid

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Allocate to</th>
<th>Insert rule</th>
<th>Sort</th>
<th>Remove</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Provider is under investigation</td>
<td>Suspected Provider</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Anomaly is detected for the claim</td>
<td>Anomaly Claim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Absolute or relative high submitted charge</td>
<td>High Charge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 High fraud rate service and more frequent visit</td>
<td>High Fraud Rate Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Remainder</td>
<td>Mixed Risks</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## WhatIf?

### Simulation Data Source
- FSP Database source

### Simulation Date
- From: 2012-06-07 01:53:21
- To: 

### Matrix Settings
#### Claim Area
<table>
<thead>
<tr>
<th>Health Care Cla</th>
<th>Model actions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspend</td>
</tr>
<tr>
<td>Suspend</td>
<td></td>
</tr>
<tr>
<td>Hold</td>
<td>Suspend</td>
</tr>
<tr>
<td>Approve</td>
<td>Suspend</td>
</tr>
</tbody>
</table>

#### Risk Factor

### Report Settings
- Name: Run 2

### WhatIf Results
- View: All results, All runs

<table>
<thead>
<tr>
<th>Display Run</th>
<th>Action</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run1</td>
<td>Approve</td>
<td>389</td>
<td>80.5</td>
</tr>
<tr>
<td></td>
<td>Hold</td>
<td>53</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Suspend</td>
<td>41</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Total Count: 483

Total Records: 500
Maximizing Return and Minimizing Cost with the Decision Management Systems

MAJOR CHALLENGES
Major Challenge: **Volume of Data**

Marriage of Predictive Analytics and Specialized, Tuned Appliances

- SQL Pushback + In Database Mining
  - Decision Trees, K-Means, Principal Component Analysis, Regression, Tree, Bayesian Networks, Naïve Bayes Classifier, K-Nearest Neighbor, Divisive Clustering
- UDF Capability in Modeler to run virtually all SPSS analytics in Netezza

Recent performance testing results using SQL Pushback and SPSS algorithms:
- Time to Score 100M records with 10 predictors and 1 model: *4 seconds*
- Time to Score 100M records with 20 predictors and 20 models: *10 seconds*
Major Challenge: Variety of Data

1. Analyze Patterns: Opinions and Related Topics
2. Visualize Results
3. Go Back to Customer Feedback
4. Score and Use in a Predictive Model
Major Challenge: Veracity of Data is addressed with approaches like Entity Resolution & Context Accumulation

What is entity analytics?
- An entity could be an individual, vehicle, vessel etc
- Entity analytics enables an organization to resolve like entities, even when the entities do not share key values (eg ID number)
- The data can come from multiple sources or just one source
- The matching technique enables even the weakest connections to be discovered
- The result is more accurate analytics, based on correctly resolved entities.

Where would you use it?
- Where data quality and model accuracy are critical
- Financial services such as banking and insurance
- Border/National security or Customs
- Policing
### Entity Analytics – specific example

<table>
<thead>
<tr>
<th>Entity 102</th>
<th>Entity 343</th>
<th>Entity 642</th>
<th>Resolved Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td>Beth L. Johns - Parker</td>
<td>Full Liz Johns</td>
<td>Elizabeth Lisa Johns</td>
</tr>
<tr>
<td></td>
<td>BL Johns</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Addr1</strong></td>
<td>123 Main Street 777 Park Road</td>
<td>33 Red Dr</td>
<td>33 Reed Dr</td>
</tr>
<tr>
<td><strong>City</strong></td>
<td>New York</td>
<td>Mamaroneck</td>
<td>White Plains</td>
</tr>
<tr>
<td><strong>State</strong></td>
<td>NY</td>
<td></td>
<td>NY</td>
</tr>
<tr>
<td><strong>Phone</strong></td>
<td>212-733-1234</td>
<td>914-698-2234</td>
<td>914-698-2234</td>
</tr>
<tr>
<td><strong>DOB</strong></td>
<td>6/21/1954</td>
<td>$9,000</td>
<td>6/21/1954</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>$8,000</td>
<td>$6,000</td>
<td>$31,000</td>
</tr>
<tr>
<td><strong>Credit Debt</strong></td>
<td>$5,359</td>
<td>$3,000</td>
<td>$1,362</td>
</tr>
<tr>
<td><strong>Other Debt</strong></td>
<td>$2,009</td>
<td>$4,001</td>
<td>$4,001</td>
</tr>
<tr>
<td><strong>Debt to Income</strong></td>
<td>92.1</td>
<td>100</td>
<td>17.3</td>
</tr>
<tr>
<td><strong>Prev Default?</strong></td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td><strong>Pending Loan</strong></td>
<td>False</td>
<td>False</td>
<td>True</td>
</tr>
</tbody>
</table>

| **Name**  | Elizabeth Lisa Johns | Liz Johns |
| **Addr1** | 33 Red Dr | 33 Reed Dr |
| **City**  | Mamaroneck | White Plains |
| **State** | NY | NY |
| **Postal** | 10354 | 10354 |
| **Phone** | 914-698-2234 | 914-698-2234 |
| **DOB**   | 6/21/1954 | 6/21/1954 |
| **Income** | $48,000 | $12,722 |
| **Credit Debt** | $12,722 | $9,009 |
| **Other Debt** | $9,009 | $4,001 |
| **Debt to Income** | 113.5 | 17.3 |
| **Prev Default?** | True | False |
| **Pending Loan** | True | False | True |
Velocity of Data requires special scoring approaches

- Enable Predictive Analytics to be deployed in new domains based on streaming data