

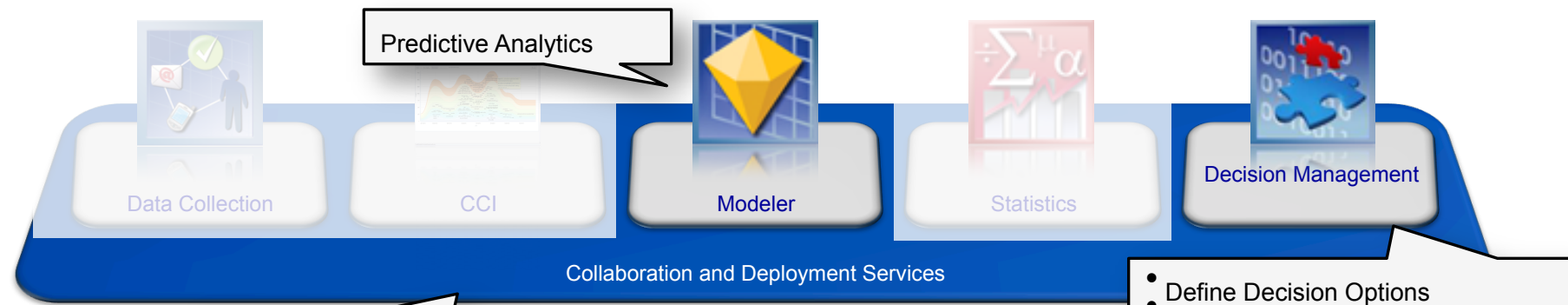
Maximizing Return and Minimizing Cost with the Decision Management Systems



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

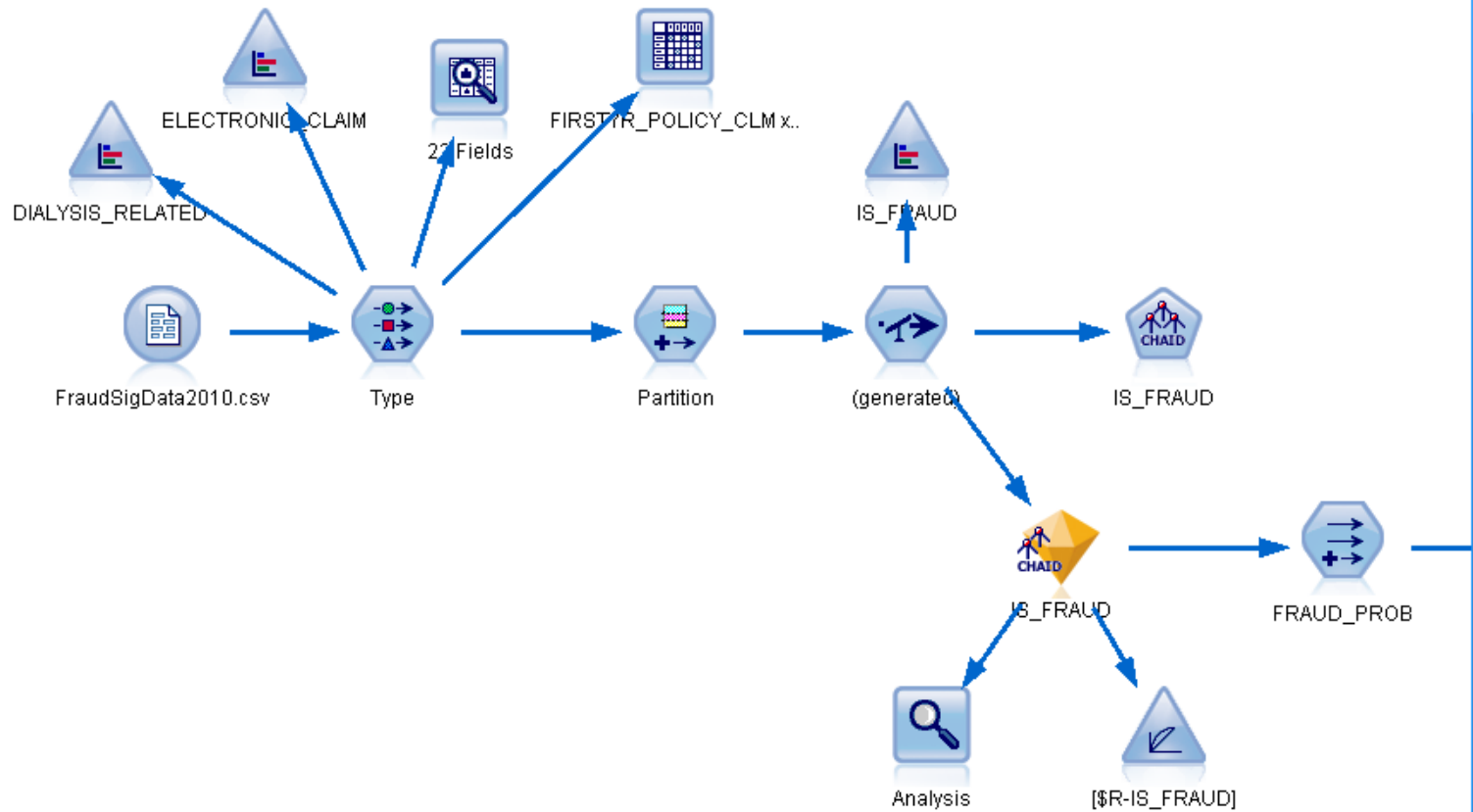
Operational Cost Take-out Agendas
Drive operational performance
Manage day to day operations
Plan for operational success



- Scoring
- Automation
- Asset Management

- Define Decision Options
- Combine Models and Rules
- What If Simulation
- Optimization



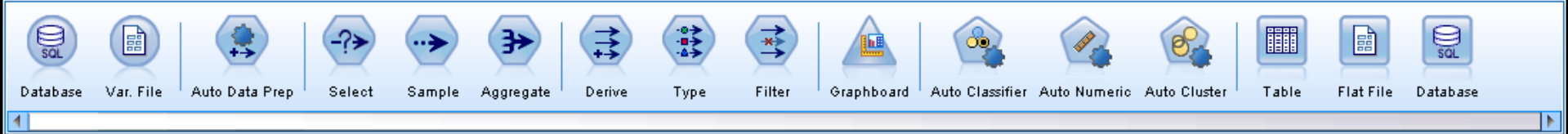


Streams Outputs Models

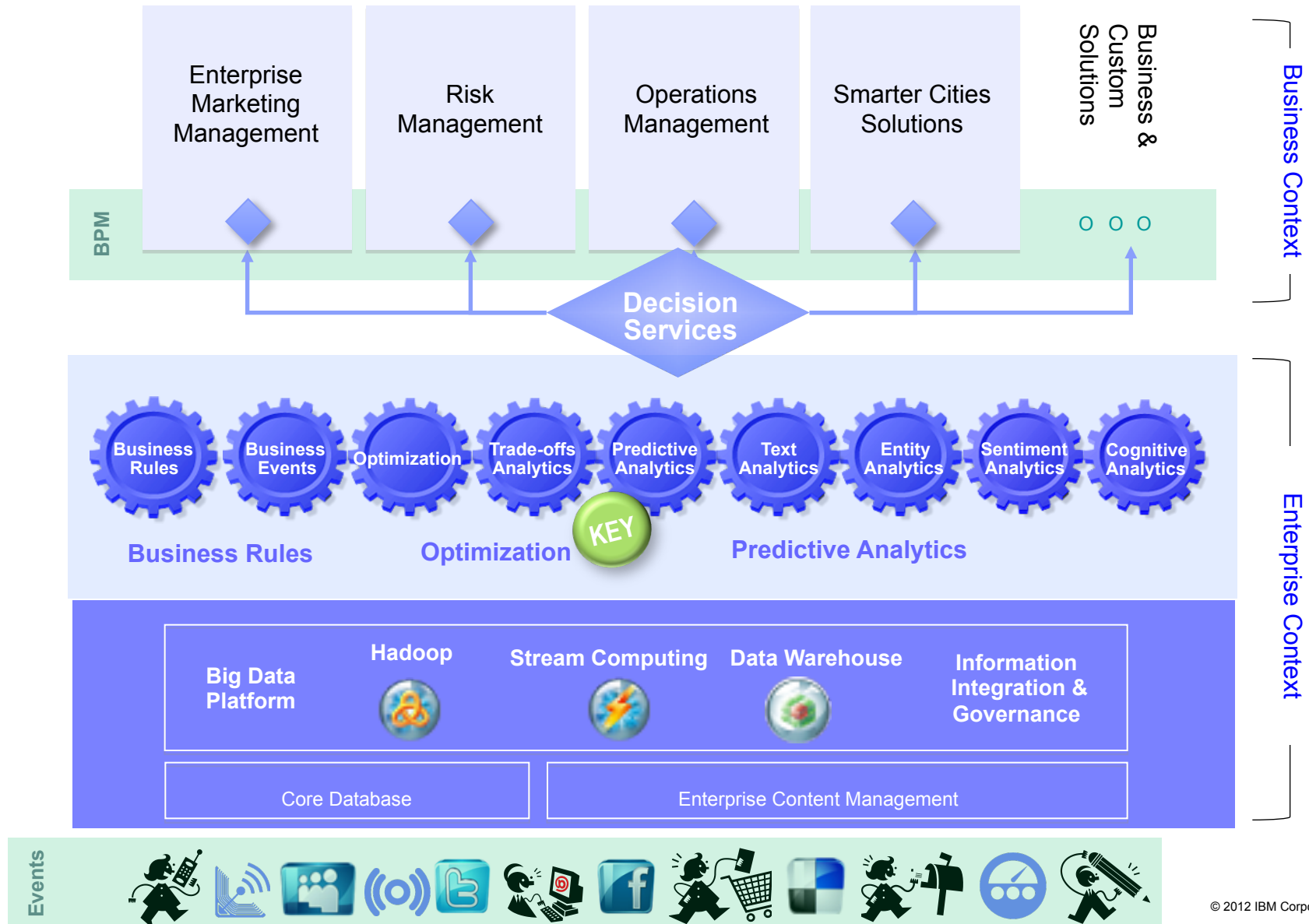
FraudSigPrototype:0

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment



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

Structured, Unstructured, Social Media. & Business Intelligence Data

Rules

Predictive Analytics

Simulation & Optimization

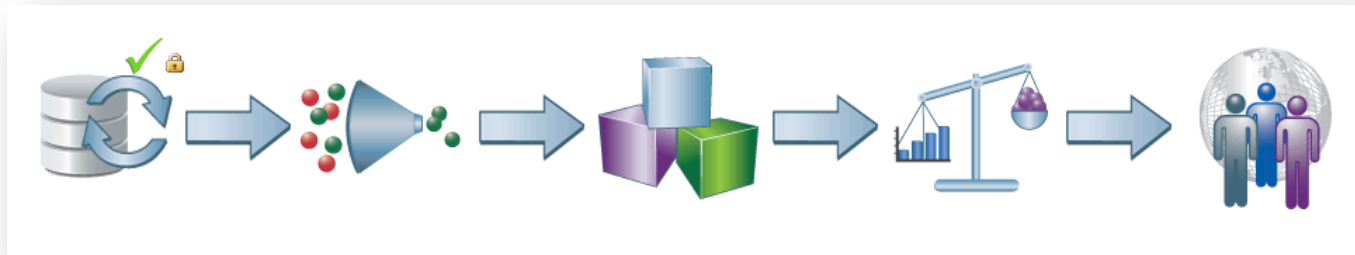
Level	Points
Low Risk	> -5
Medium Risk	> +1
High Risk	> +8

Scoring



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 Global Selections Define Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Data tab from non-administrators

Lock all Data options

Project Data Model

Data Source Fields

Data source: FSP Database source Refresh data scan Key:

<input checked="" type="checkbox"/> Operational	Field name	Measurement	Values
<input checked="" type="checkbox"/>	PROCEDURE_RISK_GROUP	Nominal	HIGH, LOW, MEDIUM, SUPER_HIGH
<input checked="" type="checkbox"/>	QUANTITY_INDEX	Continuous	[30,240]
<input checked="" type="checkbox"/>	SERVICE_TYPE	Nominal	Anesthesiology, Pharmacy, Radiology
<input checked="" type="checkbox"/>	SUBMIT_CHG	Continuous	[0,500]
<input checked="" type="checkbox"/>	SUBMITTED_CHG_INDEX	Continuous	[30,249]

Additional Fields

Add/Edit additional fields and tables

Data Sources

Project Data Sources

Add a data source

Name	No. of records	Preview	Compatible	Copy	Remove	Lock
<u>FSP Database source</u>	...		Project Data Model			

My Data Sources

Add a data source

Name	No. of records	Preview	Copy	Remove
------	----------------	---------	------	--------

Data Global Selections Define Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Global Selections tab from non-administrators

Lock all Global Selections options

Manage Global Selections



Rule name	Include/Exclude	Remove	Lock
1 Exclude_OIG_List	Exclude		

Data Global Selections **Define** Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Interaction Points



Search Dimensions:

Health Care Claim

- Approve
- Hold
- Suspend

- ▶ Health Care Claim Properties
- ▶ Choose Who This Claim Area Applies to
- ▶ Use Rules to Decide Which Action is Triggered
- ▶ Use a Model to Decide Which Action is Triggered

Simulate

Test

Data | **Global Selections** | **Define** | **Combine** | **Deploy**

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define optio

Interaction Points

Claim Area | **Risk Factor**



Search Dimensions:

Health Care Claim

Health Care Claim Properties

Simulate

Choose Who This Claim Area Applies to

Test

Use Rules to Decide Which Action is Triggered

Remainder: Apply when no rules hit

Rule name	Risk points	Sort	Remove
7 High ratio of submitted charges amount vs. average for sar	100	▲▼	✖
8 Low ratio of net eligible amount / submitted charge amount	50	▲▼	✖
9 Super high fraud rate for previous year for the procedure	50	▲▼	✖
10 High fraud rate for previous year for the procedure	20	▲▼	✖
11 Remainder	0		

Sum of Points >=	Allocate to	Remove
1 400	Suspend	✖
2 300	Hold	✖
3 0	Approve	

Use a Model to Decide Which Action is Triggered

Data Global Selections **Define** Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...

Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Search Dimensions:

Health Care Claim

Health Care Claim Properties

Choose Who This Claim Area Applies to

Use Rules to Decide Which Action is Triggered

Remainder: Apply when no rules hit

Rule name	Risk points	Sort	Remove
1 Provider is under investigation	300	▲▼	✖
2 High ratio of procedure quantity per visit vs. average for sai	50	▲▼	✖
3 Anomaly is detected for the claim	100	▲▼	✖

Interaction Points

Simulate

Test

Edit Rule

Name: Anomaly is detected for the claim

Risk points: 100

Description:

Expressions

OR Split OR

1 Fraud_AnomalyDetect_Deploy = T

2 Choose value...

3

Use

OK Cancel

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Interaction Points



Search Dimensions:

Health Care Claim

- Approve
- Hold
- Suspend

Health Care Claim Properties

- ▶ Choose Who This Claim Area Applies to
- ▶ Use Rules to Decide Which Action is Triggered
- ▼ Use a Model to Decide Which Action is Triggered

Model	Target	Measure	Delete
Fraud_Pred_Deploy.str	IS_FRAUD	Specify field...	✖

Field (FRAUD_PROB) >=	Allocate to	Remove
1 900	Suspend	✖
2 500	Hold	✖
3 Remainder	Approve	

Simulate

Test

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area



Search Dimensions:

Health Care Cla

- Approve
- Hold
- Suspend

Model - Fraud_Pred_Deploy

Model type
Predictive model [Change Model...](#)

Data source: SERVICE_TRAINING Target: IS_FRAUD

Records used to build: ...

Optional Settings

Predictive Model Results

10%
Building in Progress...

[Stop Building](#) [Build Model](#)

[Use Model](#) [Exit](#)

[Simulate](#)
[Test](#)

[Stop Building](#)
[Build Model](#)

Data Global Selections Define Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area

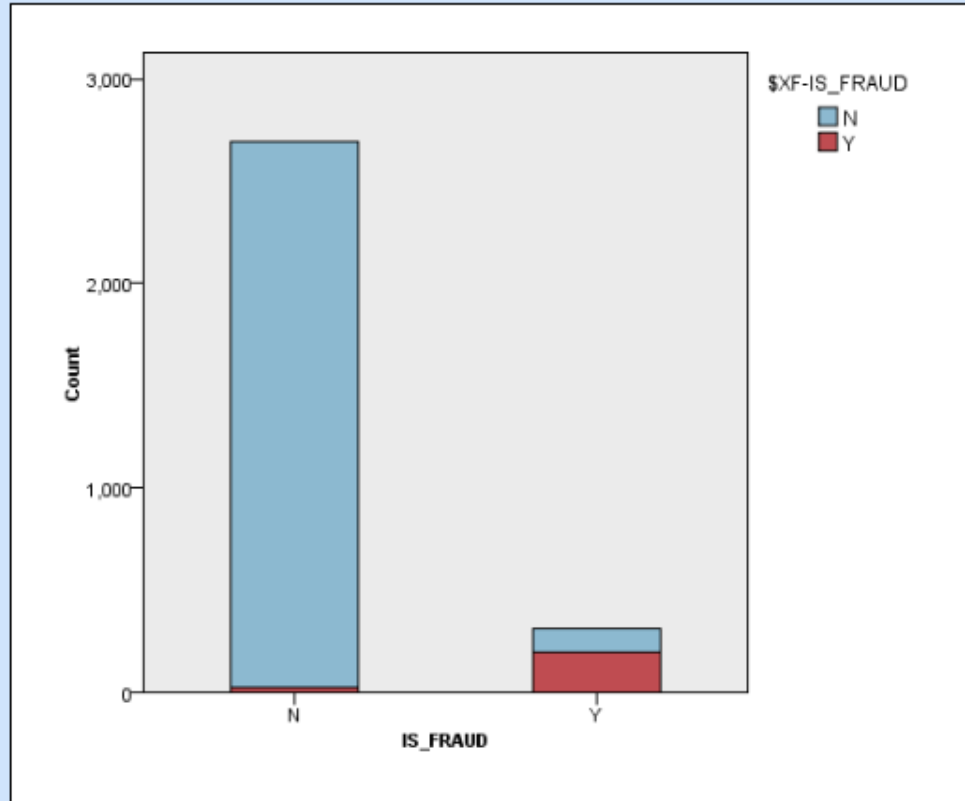
Model - Fraud_Pred_Deploy



Search Dimensions:

Health Care Cla

- Approve
- Hold
- Suspend



Distribution

Distribution

Predictor Importance

Simulate
Test

Use Model Exit

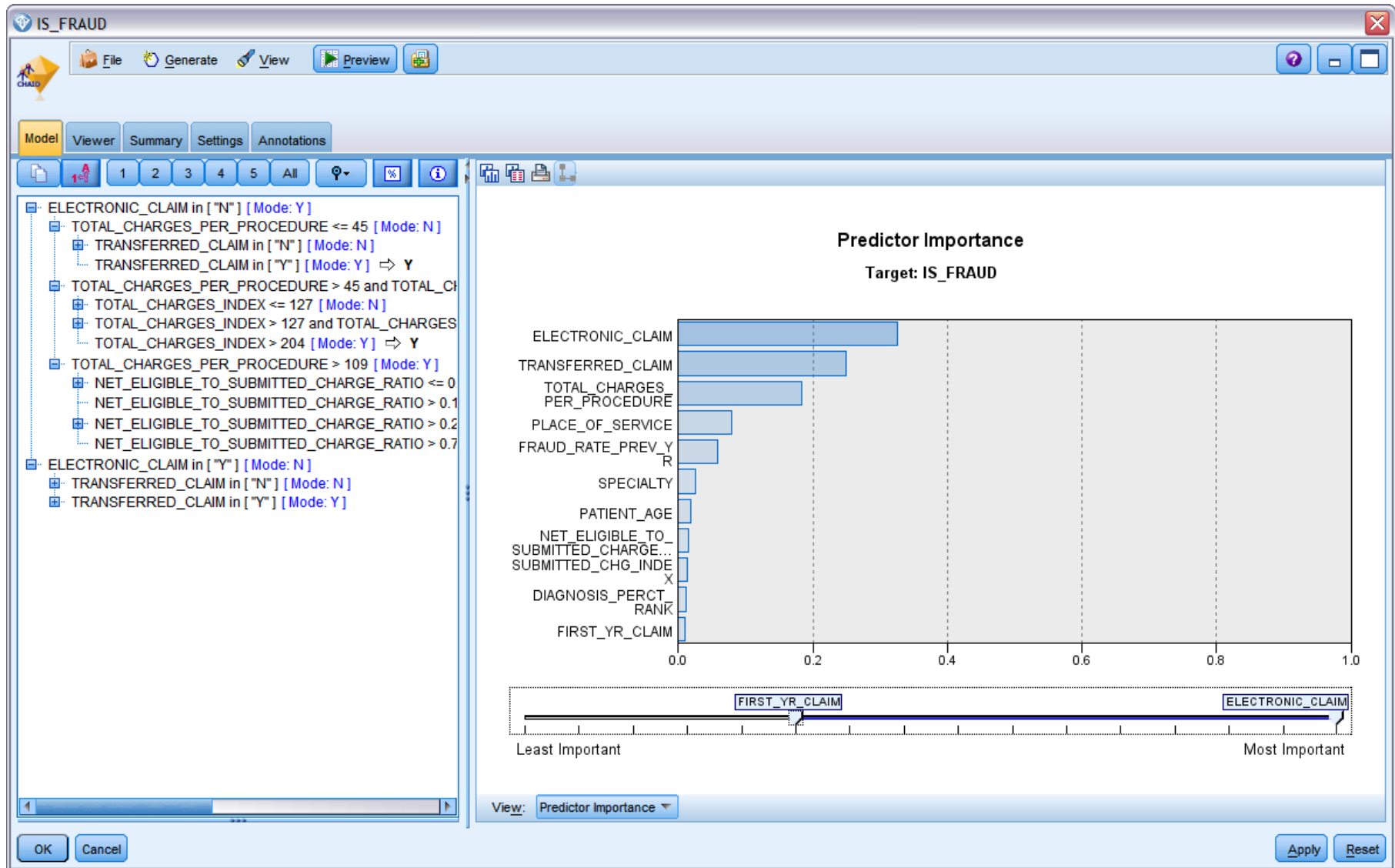
The screenshot displays the IBM SPSS Modeler interface for a project named "Fraud_Pred_Deploy:1*". The main workspace shows a workflow diagram with the following steps:

- MART.SERVICE_OIG_FIL...** (Database icon)
- IS_FRAUD** (CHAO icon, highlighted with a dashed box)
- FRAUD_PROB** (Aggregate icon)
- Filter** (Filter icon)
- Type** (Type icon)
- FRAUD_PROB** (Table icon)

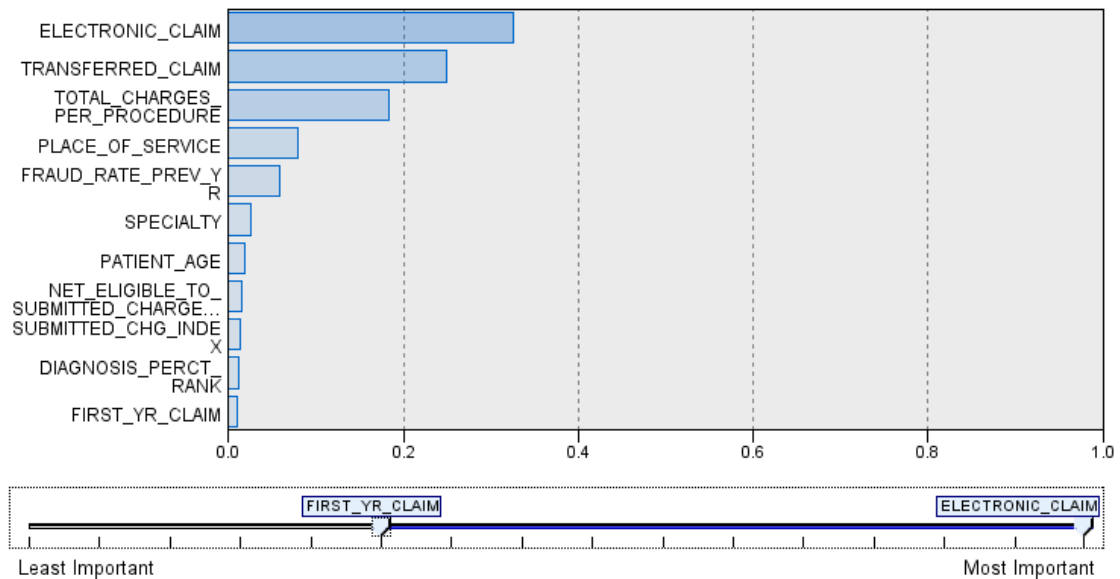
A blue arrow points from the "Type" node to a "Table" icon above the final "FRAUD_PROB" node. The right-hand pane shows the "Streams" tab with "Fraud_AnomalyDetect" and "Fraud_Pred_Deploy:1" listed. Below it, the "CRISP-DM" tab shows a project tree structure:

- (unsaved project)
 - Business Under
 - Data Understandin
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

The bottom toolbar includes various tool icons such as Database, Var. File, Auto Data Prep, Select, Sample, Aggregate, Derive, Type, Filter, Graphboard, Auto Classifier, Auto Numeric, Auto Cluster, Table, Flat File, and Database. The status bar at the bottom indicates "Server: Local Server" and "227MB / 375MB".



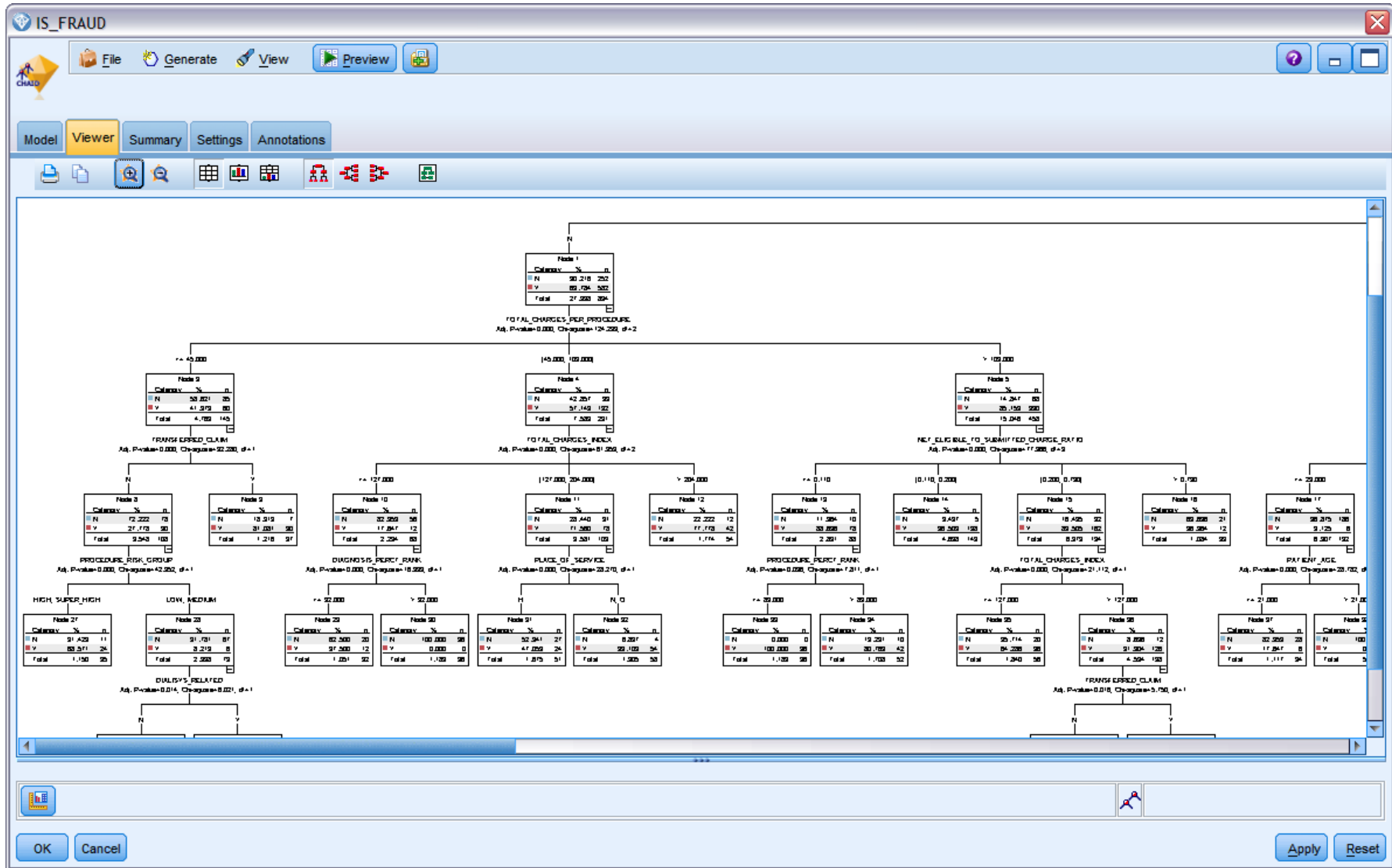
Predictor Importance
Target: IS_FRAUD



View: Predictor Importance

OK Cancel

Apply Reset



lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Interaction Points



Simulate

Dimensions:

Test

- ▶ 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					
	Rule name	Allocate to	Insert rule	Sort	Remove
1	Provider is under investigation	Suspected Provider			
2	Anomaly is detected for the claim	Anomaly Claim			
3	Absolute or relative high submitted charge	High Charge			
4	High fraud rate service and more frequent visit	High Fraud Rate Se			
5	Remainder	Mixed Risks			

- Approved Claim
- N/A
- Old Claim
- ★ High Charge
- ★ High Fraud Rate Service
- Paper Claim
- ★ Mixed Risks
- Suspended Claim**
- Suspected Provider
- Anomaly Claim
- ★ High Charge
- ★ High Fraud Rate Service
- ★ Mixed Risks

Data Global Selections Define Combine Deploy

Lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Interaction Points



- Suspended Claim Properties
- Choose Who This Risk Factor Applies to
- Allocate Using Segment Rules

Simulate Test

Search Dimensions:

- Approved Claim
- Held Claim
 - High Charge
 - High Frequency
 - High Fraud Rate Service
 - Paper Claim
 - Mixed Risks
- Suspended Claim
 - Suspected Provider
 - Anomaly Claim
 - High Frequency
 - High Charge
 - High Fraud Rate Service
 - Mixed Risks

Edit Rule

Name: Absolute or relative high charge

Allocation: High Charge

Description:

Expressions

OR Split OR

<input checked="" type="checkbox"/>	ClusterModel_MA v2	=	High Charge
<input checked="" type="checkbox"/>	SUBMITTED_CHG_INDEX	>=	200
<input checked="" type="checkbox"/>	SUBMIT_CHG	>=	400
<input type="checkbox"/>	Choose value...		

OK Cancel

The screenshot displays the IBM SPSS Modeler interface with a workflow and two dialog windows.

Workflow: The main workspace shows a workflow starting with 'SERVICE_TRAINING' (SQL) leading to 'CustomPostDataSet', 'type', 'autodataprep', 'partition', and 'cluster'. A second path starts with 'SERVICE_TRAINING' leading to 'CustomPostDataSet-Co..', 'type', 'autodataprep', 'cluster', '\$XC-autocluster', and 'Custo'.

cluster dialog (top right): This dialog is in the 'Expert' tab. It shows 'Estimated number of models to be executed: 3'. The 'Models used' table is as follows:

Use?	Model type	Model parameters	No of models
<input checked="" type="checkbox"/>	Kohonen	Default	1
<input checked="" type="checkbox"/>	K-means	Default	1
<input checked="" type="checkbox"/>	TwoStep	Default	1

Additional options include 'Restrict maximum time spent building a single model to 15 minutes' and 'Stopping rules...'. Buttons for 'OK', 'Run', 'Cancel', 'Apply', and 'Reset' are present.

cluster dialog (bottom): This dialog is in the 'Model' tab. It shows a table of generated models:

Us...	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters	Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/Largest	Importance
<input checked="" type="checkbox"/>		K-means 1	< 1	0.305	5	284	9	814	27	0.349	0

Buttons for 'OK', 'Cancel', 'Apply', and 'Reset' are present.

lock project (other users will be unable to edit)

Fraud Screening and Preve...



Hide Define tab from non-administrators

Lock all Define options

Claim Area Risk Factor

Interaction Points



- ▶ Suspended Claim Properties
- ▶ Choose Who This Risk Factor Applies to
- ▼ Allocate Using Segment Rules

Simulate

Test

Allocate using rules Multiple Allocation Allocate randomly

Allocate to: First valid

	Rule name	Allocate to	Insert rule	Sort	Remove
1	Provider is under investigation	Suspected Provider			
2	Anomaly is detected for the claim	Anomaly Claim			
3	Absolute or relative high submitted charge	High Charge			
4	High fraud rate service and more frequent visit	High Fraud Rate Se			
5	Remainder	Mixed Risks			

Dimensions:

- Approved Claim
- N/A
- Old Claim
- ★ High Charge
- ★ High Fraud Rate Service
- Paper Claim
- ★ Mixed Risks
- Suspended Claim**
- Suspected Provider
- Anomaly Claim
- ★ High Charge
- ★ High Fraud Rate Service
- ★ Mixed Risks

Simulation

Simulation Data Source

FSP Database source

Simulation Date

From: 2012-06-07 01:52:15 No expiration

To: 15

Output Options

Run

Claim Area

Risk Factor

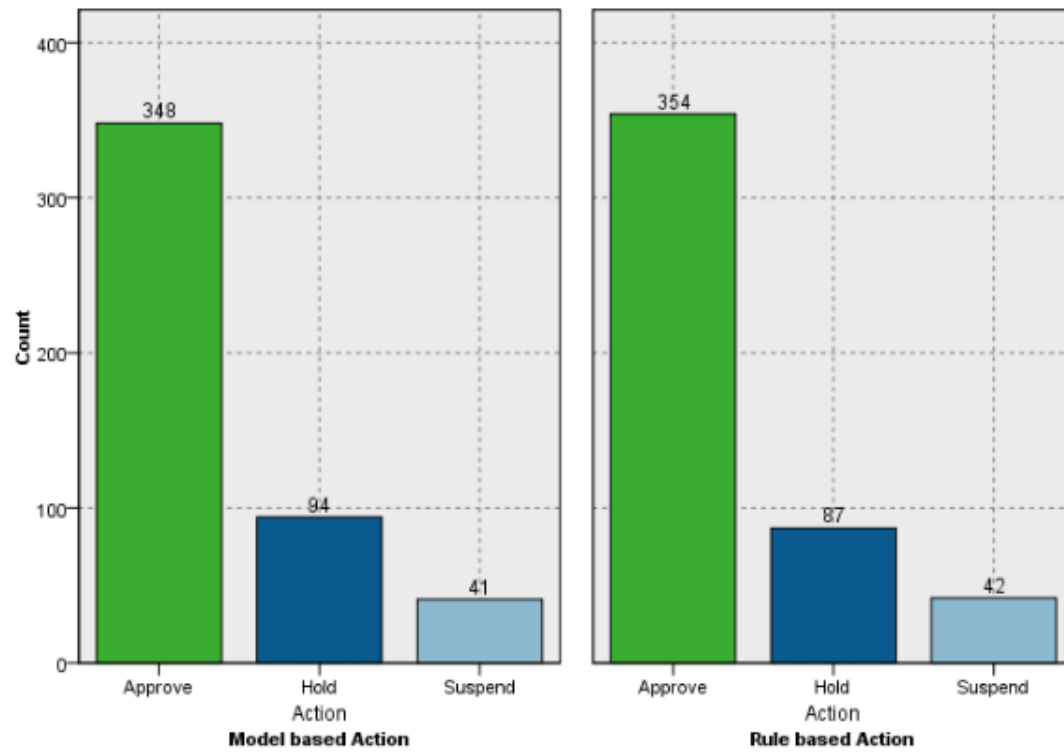
- Claim Area
- Action
- Rule

Graph

Data source: FSP Database source

Total Simulation Records: 500

View: Count



Action - Final Outcome:

WhatIf?

Simulation Data Source

FSP Database source

Simulation Date

From: 2012-06-07 01:53:21 No expiration

To:

Matrix Settings

Claim Area

Risk Factor

Health Care Cla

		Model actions		
		Suspend	Hold	Approve
Rules actions	Suspend	Suspend	Hold	Hold
	Hold	Suspend	Hold	Approve
	Approve	Suspend	Approve	Approve

Report Settings

Name: Run 2

WhatIf Results

View All results All runs

Claim Area

Risk Factor

Display Run1

Number of runs retained: 1

Action	Count	Percent
Approve	389	80.5
Hold	53	11
Suspend	41	8.5

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

2. Visualize Results

1. Analyze Patterns: Opinions and Related Topics

Global	In	Type 1	Type 2
148		<Unknown>	
66		<Products>	
64		<Positive>	
38		<Negative>	
37		<Unknown>	<Positive>
33	fx	<CustomerSupport>	<Positive>
29		<Unknown>	<Negative>
28		<CustomerSupport>	
20		<Unknown>	<Contextual>
19	fx	<PositiveAttitude>	
14		<Products>	<Positive>

4. Score and Use in a Predictive Model

3. Go Back to Customer Feedback

Global	Docs	In	Concept 1	Concept 2
2		2	fx car	out of stock
1		1	fx car	missed
1		1	fx car	bad
1		1	fx car	wrong
1		1	fx car rental company	no intention to use
1		1	fx vacation car rental	problem
1		1	fx windshield wipers	problem
1		1	fx car	problem
1		1	fx car	too long
1		1	fx damage to the car	problem

Customer_Service (12)	Categories
1 They were idiots. The car had problems and they were unable to fix them or provide a replacement without a lot of hassle.	Neg: Product: Functioning Neg: Service: Knowledge Pos: Service: General
2 I thought the representative handled the initial situation badly. The company was out of cars, with none coming in that day. Then the representative tried to find us a car at another franchise. There they were successful.	Neg: Product: Availability/Varie... Other: Don't Know Other: No Like/No Dislike
3 The car rental company that I went with had very good customer service. They were out of a certain car I reserved and gave me a upgrade and apologized.	Pos: Service: Attitude Pos: Service: General Neg: Product: Availability/Varie...
Was average, nothing out of the ordinary. Took a long time to get the car.	Neg: Product: Functioning Neg: Service: Accessibility

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

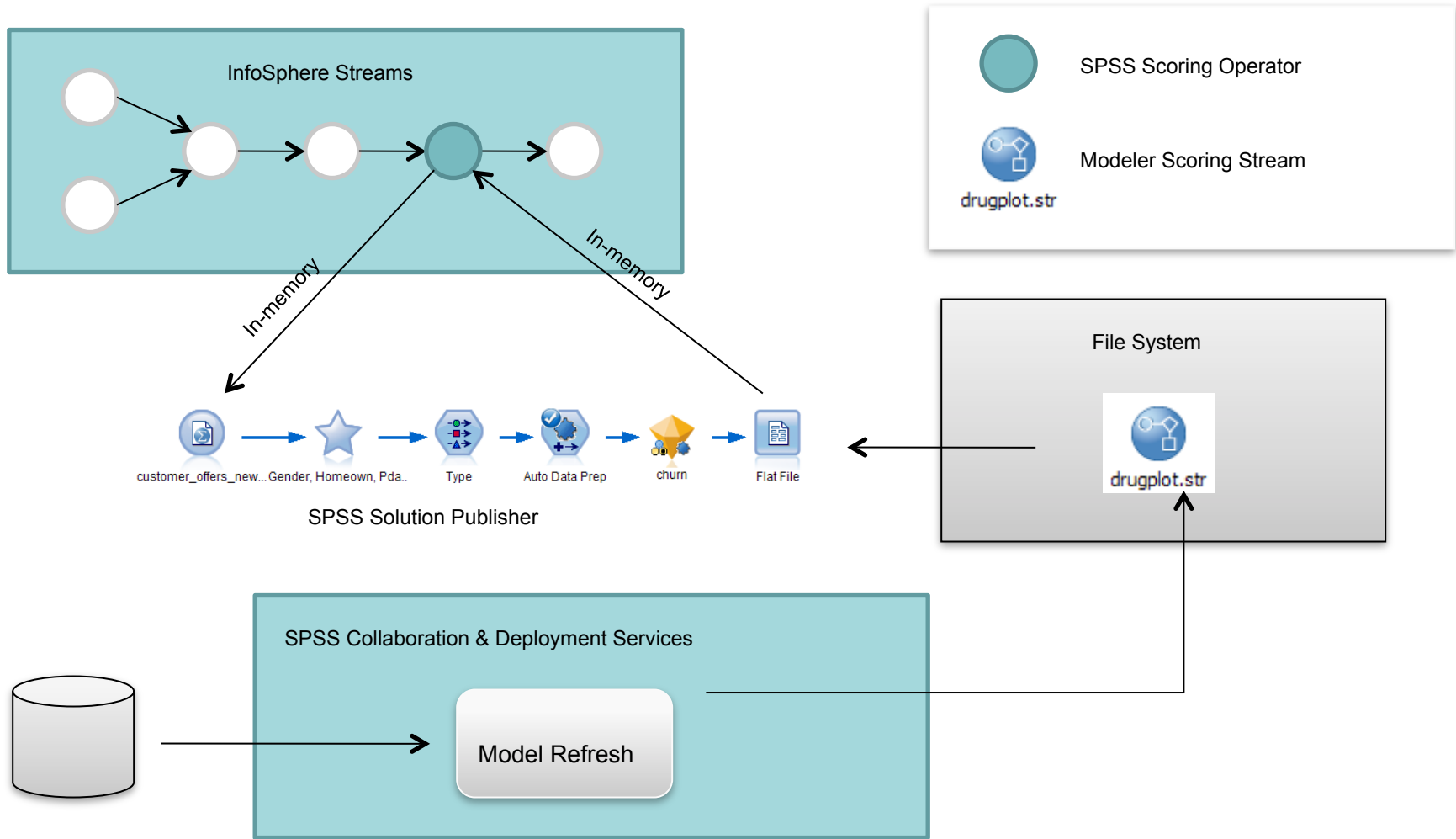


Entity Analytics – specific example

Entity 102	Entity 343	Entity 642	Resolved Entity
Name Beth L. Johns - Parker BL Johns Addr1 123 Main Street 777 Park Road City New York State NY Phone 2127331234 DOB 6/21/1954 Income \$8,000 Credit Debt \$5,359 Other Debt \$2,009 Debt to Income 92.1 Prev Default? True Pending Loan False	Full Liz Johns Addr1 33 Red Dr City Mamaroneck State NY Postal 10354 Phone 212-733-1234 Income \$9,000 Credit Debt \$6,000 Other Debt \$3,000 Debt to Income 100 Prev Default? True Pending Loan False	Full Elizabeth Lisa Johns Addr1 33 Reed Dr City White Plains State NY Postal 10354 Phone 914-698-2234 Income \$31,000 DOB 6/21/1954 Credit Debt \$1,362 Other Debt \$4,001 Debt to Income 17.3 Prev Default? False Pending Loan True	Name Elizabeth Lisa Johns Liz Johns Beth L Johns-Parker BL Johns Addr1 123 Main Street 777 Park Road 33 Red Dr 33 Reed Dr City New York, White Plains, Mamaroneck, NY Postal 11732, 10354 Phone 212-733-1234 914-698-2234 DOB 6/21/1954 Defaults Yes Income \$48,000 Credit Debt \$12,722 Other Debt \$9,009 Debt to Income 113.5 Prev Default? True Pending Loan True

Velocity of Data requires special scoring approaches

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



IBM®