

IBM Software



# Business Analytics **Forum** 2010

See The Future Of Decision Making

**November 3-5, 2010**

Gold Coast – QLD – Australia

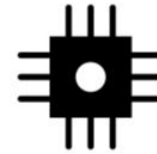
## The Power of Data Mining

Brad Hill  
IBM Business Analytics





# The world is changing, enabling organizations to make faster, better-informed decisions



**Instrumented**



**Interconnected**



**Intelligent**





**With this change comes an explosion in information ...**



Volume of Digital Data



Variety of Information



Velocity of Decision Making

**... Yet organizations are operating with blind spots**

**Lack of Insight**

**1 in 3** managers frequently make critical decisions without the information they need

**Inefficient Access**

**1 in 2** don't have access to the information across their organization needed to do their jobs

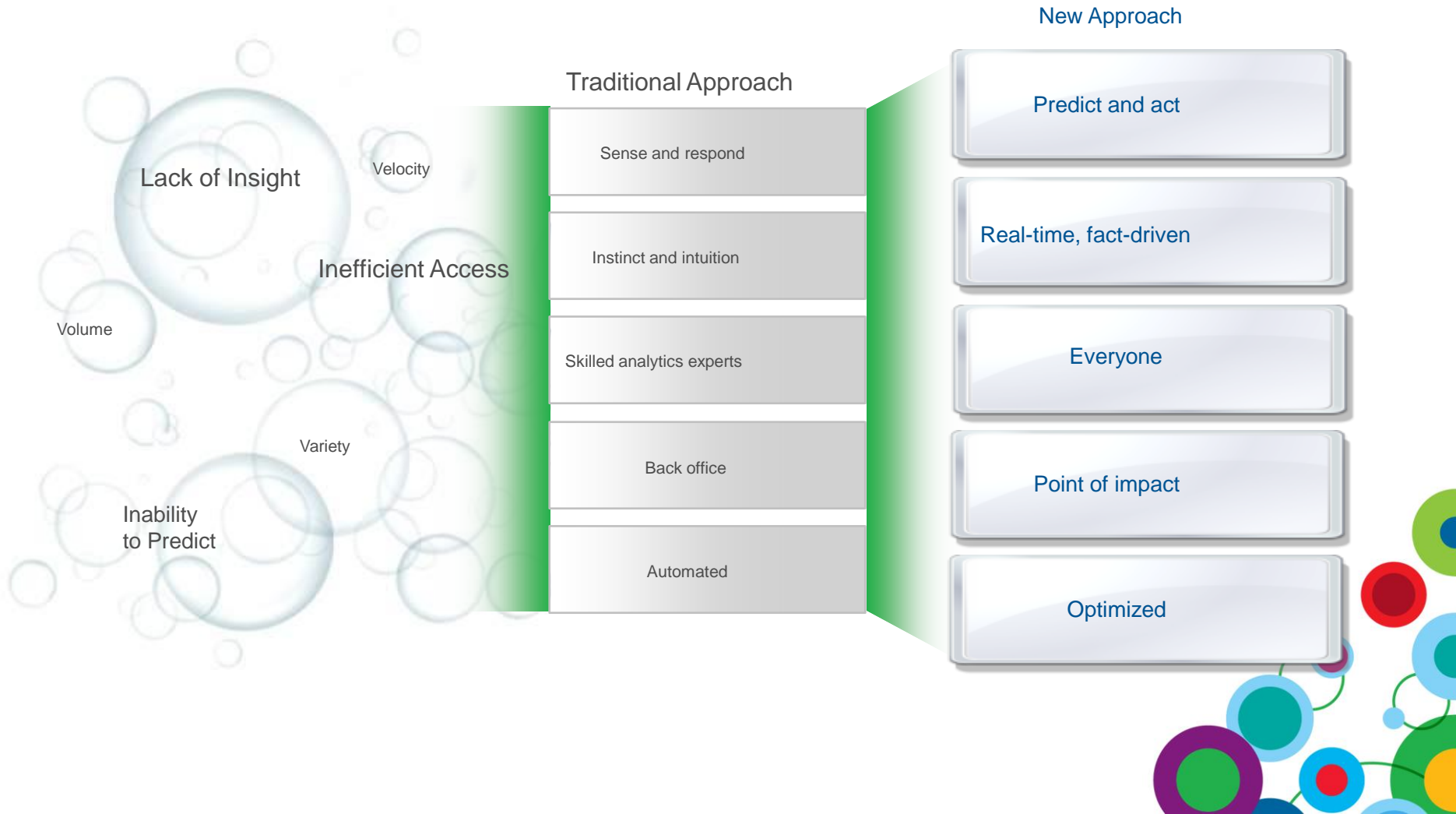
**Inability to Predict**

**3 in 4** business leaders say more predictive information would drive better decisions

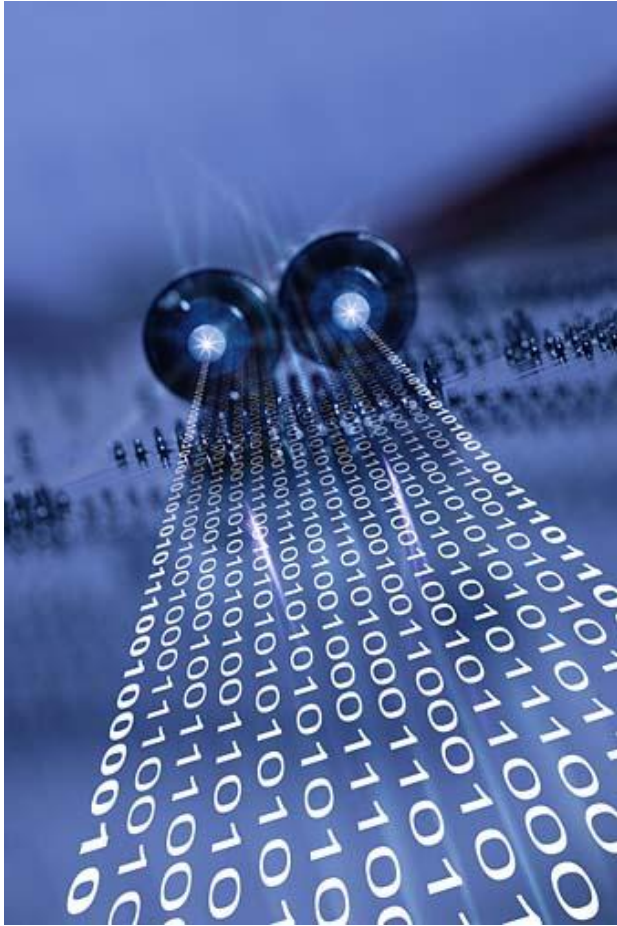




# Business challenges and conditions have placed a renewed urgency on business analytics and optimization

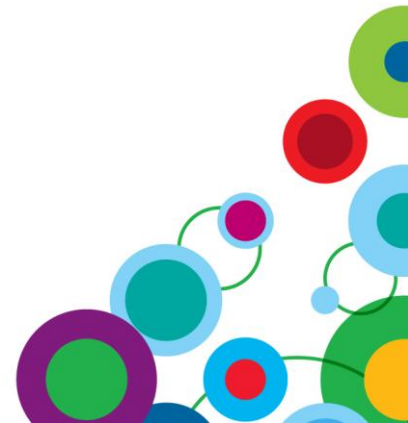


# What is predictive analytics?



Predictive Analytics helps connect **data** to effective **action** by drawing reliable conclusions about current conditions and **future** events

*Gareth Herschel, Research Director, Gartner Group*





# Predictive analytics in action

## ■ Customer relationship management

### “analytical CRM”

- Who are our best customers?
- Can we get more like that?
- What/why do they buy?
- Why do they leave?

## ■ Human capital management

- Who are our best employees?
- How do we keep our best employees from leaving?
- Which prospects should we recruit?

## ■ Science

- Genetics
- Drug discovery
- Medical research
- Food authentication

## ■ Industrial process optimisation

## ■ Fraud detection

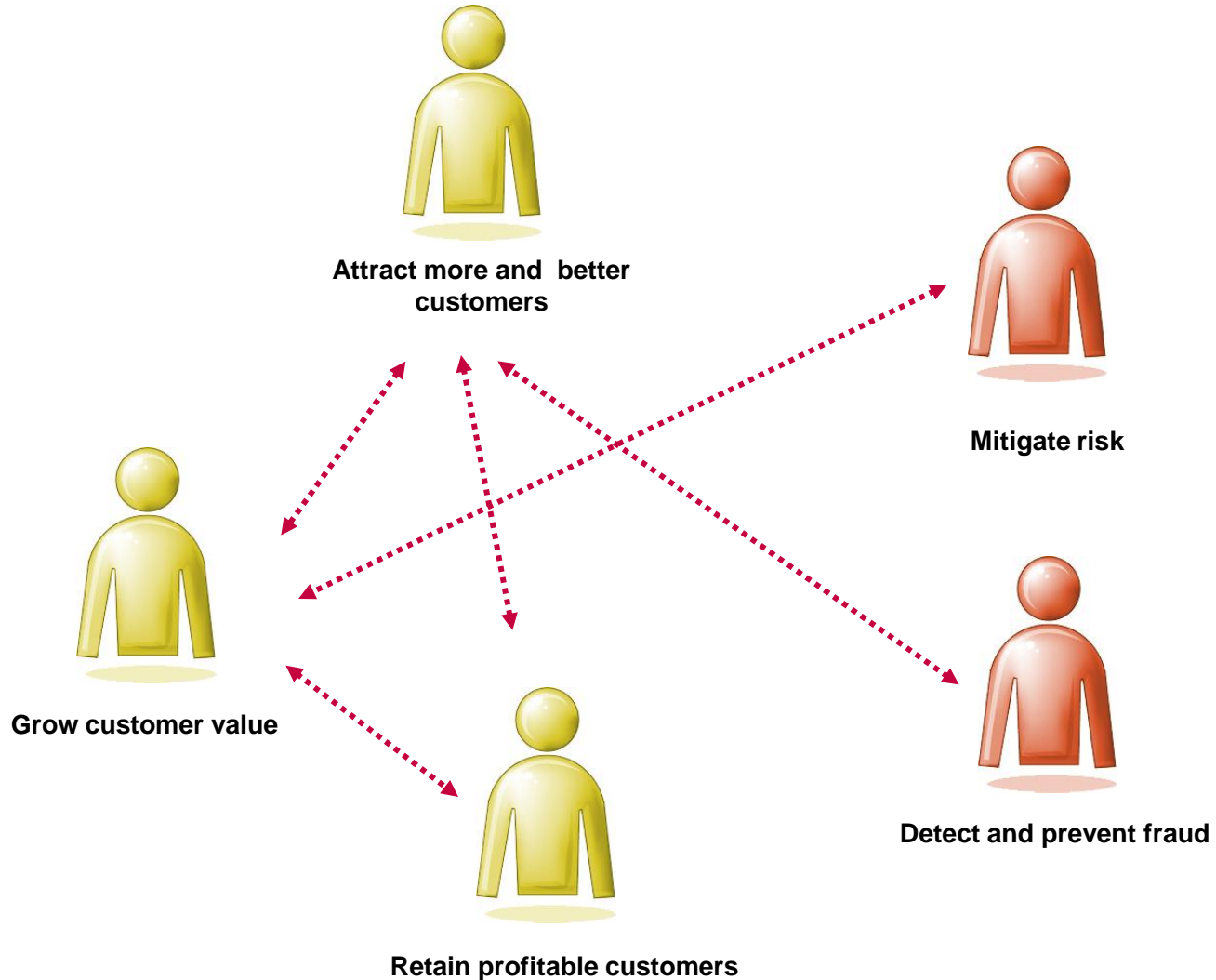
- Money laundering
- Network intrusion
- Tax audits & collection

## ■ Crime analysis

and many more...



# Business objectives





# Data mining and text analytics

## Data Mining

- Use advanced analytical techniques on data
- Discover key relationships between variables
- Model effect of variables on outcomes
- Determine influence on outcomes
- Predict outcomes
- Apply models to new data

## Text Analytics

- Extract, analyse and create structure for unstructured data
- Integrate analysis results into operational systems
- Integrate analysis results into Business Intelligence applications
- Integrate analysis results with structured data and use as input for Data Mining
- Improves model accuracy







# Complete workbench

The screenshot displays the PASW Modeler 14 interface with a workflow diagram. The workflow starts with three input nodes: 'Comments.xls', 'Customers', and 'Churn'. These are merged into a 'Comments' node. A callout box 'Join the datasets based on ID' points to the Merge node. The 'Comments' node is then processed by a 'Type' node, resulting in 'CHURN x [25 Fields]'. This is followed by another 'Type' node, which branches into two paths: one leading to a 'CHURN' node and another to a 'Churn Text' node. A callout box 'Different models with and without using text' points to these two paths. The 'Churn Text' node is processed by a 'Type' node, resulting in 'CHURN'. A callout box 'Compare the models' points to the final 'CHURN' node. The workflow concludes with an 'Analysis' node.

The interface includes a menu bar (File, Edit, Insert, View, Tools, SuperNode, Window, Help), a toolbar with various icons, and a sidebar on the right with 'Streams' and 'Outputs' tabs. The 'Streams' tab shows 'Stream1' and 'Hands on 2d'. The 'Outputs' tab shows 'CRISP-DM' and 'Classes'. The 'Classes' tab shows a project structure: '(unsaved project)' > 'Business Understanding' > 'Data Understanding' > 'Data Preparation' > 'Modeling' > 'Evaluation' > 'Deployment'. The bottom of the interface features a 'Favorites' bar and a 'Sources' bar with various tool icons like 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 shows 'Server: Local Server' and '252MB / 420MB'.



# Complete workbench

Hands on 2d\* - PASW® Modeler 14

File Edit Insert View Tools SuperNode Window Help

Comments.xls  
Customers  
Churn  
Merge  
Comments  
CHURN x [25 Fields]  
Type  
CHURN  
Type  
Churn Text  
CHURN  
Analysis  
Join the datasets based on ID  
Different models with and without using text  
Compare the models

Streams Outputs Models  
Stream1  
Hands on 2d

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

Favorites Sources Record Ops Field Ops Graphs Modeling Database Modeling Output Export PASW® Statistics PASW® Text Analytics

Database Var. File Auto Data Prep Select Sample Aggregate Derive Type Filter Graphboard Auto Classifier Auto Numeric Auto Cluster Table Flat File Database

Server: Local Server 252MB / 420MB



# Complete workbench

The screenshot displays the PASW Modeler 14 software interface. The main workspace shows a workflow diagram with the following components:

- Inputs:** Comments.xls, Customers, and Churn.
- Merge:** A central node where the inputs are combined. A callout box says "Join the datasets based on ID".
- Comments:** A node receiving data from the Merge node.
- CHURN x [25 Fields]:** A node receiving data from the Comments node.
- Type:** Two nodes receiving data from the CHURN x [25 Fields] node.
- CHURN:** Two nodes receiving data from the Type nodes. A callout box says "Different models with and without using text".
- Churn Text:** A node receiving data from the CHURN nodes.
- CHURN:** A second CHURN node receiving data from the Churn Text node.
- Analysis:** A final node receiving data from the CHURN node. A callout box says "Compare the models".

The interface includes a menu bar (File, Edit, Insert, View, Tools, SuperNode, Window, Help), a toolbar with various icons, and a sidebar on the right with "Streams" and "Outputs" tabs. The "Streams" tab shows "Stream1" and "Hands on 2d". The "Outputs" tab shows "CRISP-DM" and "Classes". The "Classes" tab shows a project structure:

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

The bottom of the interface features a "Favorites" bar and a "Sources" bar with tabs for Sources, Record Ops, Field Ops, Graphs, Modeling, Database Modeling, Output, Export, PASW® Statistics, and PASW® Text Analytics. Below this is a toolbar with icons for 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 shows "Server: Local Server" and "252MB / 420MB".



# Complete workbench

The screenshot displays the PASW Modeler 14 software interface. The main workspace contains a workflow diagram with the following components and annotations:

- Inputs:** 'Comments.xls', 'Customers', and 'Churn' data sources.
- Merge:** A central node where the inputs are combined. A yellow callout box above it says "Join the datasets based on ID".
- Comments:** A node receiving data from the Merge node.
- CHURN x [25 Fields]:** A node receiving data from the Comments node.
- Type:** Two nodes that process the CHURN x [25 Fields] data. A yellow callout box between them says "Different models with and without using text".
- CHURN:** A node receiving data from the top Type node.
- Churn Text:** A node receiving data from the bottom Type node.
- CHURN:** A second CHURN node receiving data from the Churn Text node.
- Analysis:** A final node receiving data from the second CHURN node. A yellow callout box above it says "Compare the models".

The interface includes a menu bar (File, Edit, Insert, View, Tools, SuperNode, Window, Help), a toolbar with various icons, and a right-hand sidebar with 'Streams' and 'Outputs' tabs. The 'Streams' tab shows 'Stream1' and 'Hands on 2d'. The 'Outputs' tab shows a CRISP-DM process flow: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The bottom of the interface features a 'Favorites' bar and a 'Toolbox' with icons for 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 '252MB / 420MB'.



# Complete workbench

Hands on 2d\* - PASW® Modeler 14

File Edit Insert View Tools SuperNode Window Help

Streams Outputs Models

- Stream1
  - Hands on 2d

CRISP-DM Classes

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

Comments.xls  
Customers  
Churn  
Merge  
Comments  
CHURN x [25 Fields]  
Type  
CHURN  
Type  
Churn Text  
CHURN  
Analysis  
Join the datasets based on ID  
Different models with and without using text  
Compare the models

Database Var. File Auto Data Prep Select Sample Aggregate Derive Type Filter Graphboard Auto Classifier Auto Numeric Auto Cluster Table Flat File Database

Server: Local Server 252MB / 420MB



# Data mining techniques

Technique	Algorithms	Usage
Classification (or prediction)	Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc	Used to predict group membership (ie will this employee leave?) or a number (ie how many widgets will I sell?)





# Data mining techniques

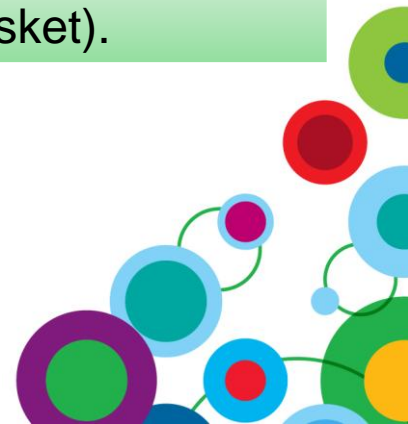
Technique	Algorithms	Usage
Classification (or prediction)	Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc	Used to predict group membership (ie will this employee leave?) or a number (ie how many widgets will I sell?)
Segmentation	Auto Clustering, K-means, etc.  Anomaly detection	Used to classify data points into groups that are internally homogenous and externally heterogeneous.  Identify cases that are unusual





# Data mining techniques

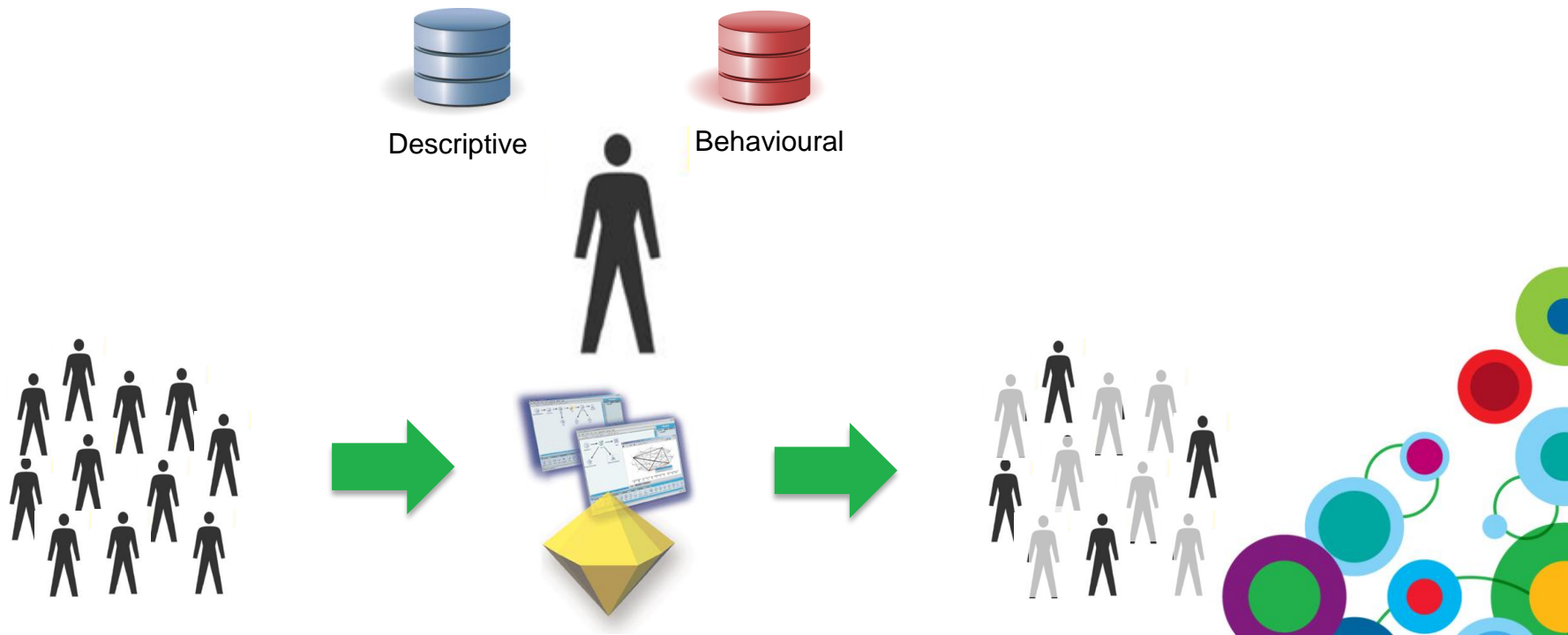
Technique	Algorithms	Usage
Classification (or prediction)	Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc	Used to predict group membership (ie will this employee leave?) or a number (ie how many widgets will I sell?)
Segmentation	Auto Clustering, K-means, etc.  Anomaly detection	Used to classify data points into groups that are internally homogenous and externally heterogeneous.  Identify cases that are unusual
Association	APRIORI, Carma, Sequence	Used to find events that occur together or in a sequence (ie market basket).





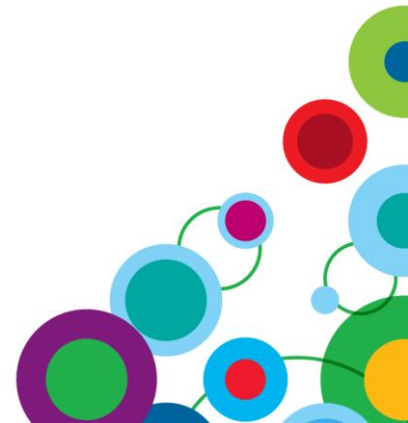
# Scenario

- Customer and product data
- Explore and understand data
- Build a model to identify customers likely to respond
- Generate a list for marketing





# Demonstration





# How predictive intelligence gets deployed

The screenshot displays a call center agent's interface. At the top, there are navigation tabs: "my activities", "team activities", "products", "charts", and "introductions". On the right, it says "Welcome, John Palmer" with "log out" and "help" links. Below this are buttons for "end call", "hold call", and "direction".

The main interface is divided into several sections:

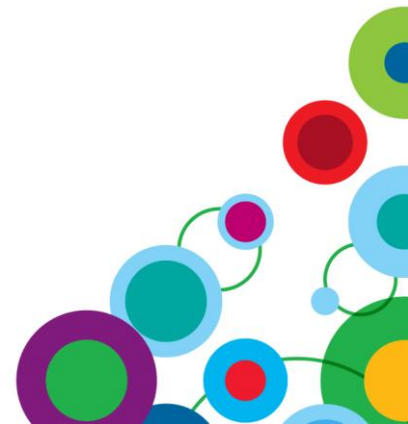
- Customer Details:** A form with fields for Last name (Nes), Gender (M), First name (Frank), Address (Crossground), City (Catburg), Zipcode (6893 OK), Age (43), and Profession (Manager). A "get info" button is at the bottom left.
- Products:** A table with columns ID, Description, and Group.

ID	Description	Group
12	Teen Visa Card	Banking
13	Home Equity Loan	Banking
14	Easy Access Account	Banking
- Contact history:** A table with columns Description, Date, and Result.
- Details current call:** A section with a dropdown menu for "Type of contact" (set to "-- To be determined --") and a "Description" text area. A "submit" button is at the bottom right.
- Recommendation:** A section with columns Interaction and Offer. It shows "Prevent Churn-HV (Single)" and "Retention - Racing". An "Action" dropdown menu is open, showing options: "-- Select reason", "F011: Conversation took too long", "F012: Customer not in the mood", and "F013: Already on target: Not in t".
- Message:** A text input field at the bottom left.

A call center agent submits customer information during an interaction

Based on the predictive model, a single offer is presented to the customer

The reaction to the offer is tracked and used to refine the model





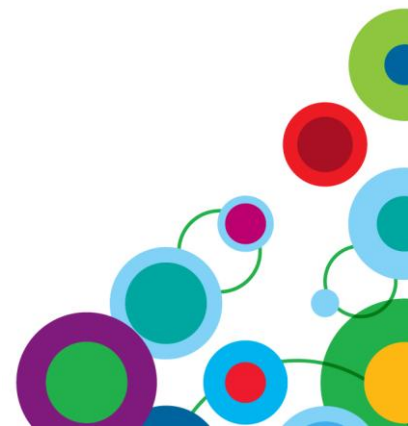
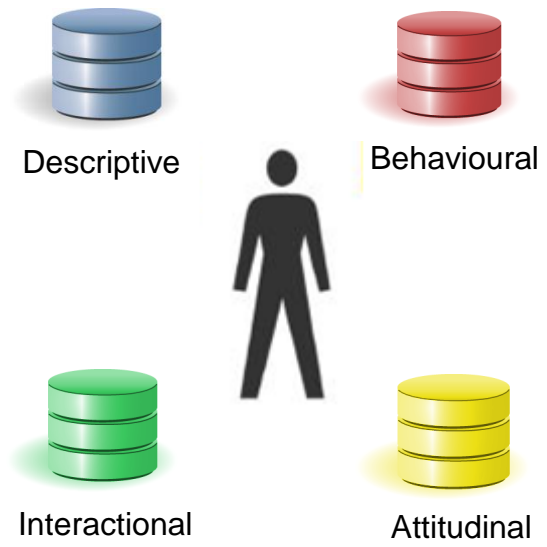
# The importance of text



Because people communicate with **words**, not numbers, it has become critical to be able to **mine text** for its **meaning** and to sort, analyse, and understand it in the same way that data has been tamed. In fact, the two basic types of information complement each other, with data supplying the “**what**” and text supplying the “**why**”.

Source IDC: “Text Analytics: Software’s Missing Piece?”







# Text mining within IBM SPSS Modeler

Interactive Workbench - Q4carcomments

File Edit View Generate Categories Tools Help

Build Extend Score Display

Category	Descriptors	Docs
All Documents	-	200
Uncategorized	-	48
No concepts extracted	-	1
Pos: General Satisfaction	26	63
Pos: Product: Functioning	16	45
Neg: Product: Functioning	15	28
Neg: Service: Accessibility	38	15
Neg: Product: Design/Features	3	12
Other: Don't Know	14	11
Pos: Pricing and Billing	3	8
Neg: Product: Availability/Variety/Size	12	6
Neg: General Dissatisfaction	14	6
Contb: Pricing and Billing	5	6
Contb: Company: Public Image/Reputation	3	6

Category Bar Category Web Category Web Table

Category	Bar	Selection %	Docs
Pos: General Satisf...	[Bar]	53.8	
Pos: Product: Functi...	[Bar]	38.5	
Neg: Product: Functi...	[Bar]	23.9	
Neg: Service: Acces...	[Bar]	6.0	
Other: Don't Know	[Bar]	4.3	
Neg: Product: Desig...	[Bar]	3.4	
Contb: Company: Pu...	[Bar]	3.4	
Pos: Pricing and Billi...	[Bar]	2.6	
Neg: General Dissat...	[Bar]	2.6	
Contb: Pricing and Bi...	[Bar]	2.6	
Pos: Service: Attitude	[Bar]	1.7	
Pos: Product: Usabil...	[Bar]	1.7	
Neg: Product: Availa...	[Bar]	1.7	
Pos: Service: General	[Bar]	0.9	
Neg: Pricing and Bill...	[Bar]	0.9	
Pos: Service: Acces...	[Bar]	0.9	

Extract Map Display

380 concepts

Concept	In	Global	Docs	Type
car	fx	106 (14%)	93 (47%)	<Products>
good	fx	57 (7%)	47 (24%)	<Positive>
clean	fx	37 (5%)	37 (19%)	<PositiveFeeling>
excellent	fx	19 (2%)	18 (9%)	<Positive>
well	fx	12 (2%)	12 (6%)	<Positive>
dirty	fx	10 (1%)	10 (5%)	<NegativeFeeling>
like	fx	9 (1%)	9 (5%)	<Positive>
dislike	fx	9 (1%)	9 (5%)	<Negative>
small	fx	8 (1%)	8 (4%)	<Contextual>
bad	fx	8 (1%)	8 (4%)	<Negative>
no problem	fx	7 (1%)	7 (4%)	<Positive>
satisfied	fx	6 (1%)	6 (3%)	<Positive>
price	fx	6 (1%)	6 (3%)	<Budget>
fun	fx	6 (1%)	6 (3%)	<Positive>
brand new	fx	6 (1%)	6 (3%)	<Positive>

Q4carcomments (117)

Doc	Text	Categories
4	Car was clean and low on miles.	Pos: Product: Functioning
5	Car was clean but the windows were very dirty on the inside.	Neg: Product: Design/Feat... Neg: Service: Accessibility Pos: Product: Functioning
6	Car was clean. It would be helpful if car contained a quick instruction sheet illustrating where things are (i.e. gas cap, trunk release) as well as how to operate the security keychain.	Pos: Product: Functioning Pos: Product: Information Pos: Service: Knowledge
7	Car was economy. Awful car. Gave us trouble. We had to change it.	Neg: Product: Functioning
8	Car was great. Air conditioning worked, it was clean, and drove really well.	Pos: Product: Functioning Pos: General Satisfaction
9	car was ok, but it had an odor from a smoker or something	Pos: General Satisfaction



# Text mining within IBM SPSS Modeler

Interactive Workbench - Q4carcomments

File Edit View Generate Categories Tools Help

Text Link Analysis

Extract 66 patterns Display

Global	In	Type1	Type2
1		<CustomerSupport>	<Negative>
1	fx	<Documentation>	<Negative>
1	fx	<Budget>	<NegativeBudget>
1	fx	<Unknown>	<NegativeBudget>
7	fx	<Products>	<NegativeFeeling>
5	fx	<Products>	<NegativeFunctioni
2	fx	<Unknown>	<NegativeFunctioni
1	fx	<Characteristics>	<NegativeFunctioni
59		<Products>	<Positive>
26		<Unknown>	<Positive>
4	fx	<Characteristics>	<Positive>
3	fx	<Budget>	<Positive>
2		<Performance>	<Positive>
1		<Location>	<Positive>
1	fx	<Documentation>	<Positive>
1		<Product>	<Positive>

Concept Web Type Web

Grid Layout

Extract Selected: 31 patterns Display

Global	Docs	In	Concept1	Concept2
3	3		condition	good
1	1	fx	speakers	blown
1	1	fx	windshield wiper	not working
1	1	fx	red car	fading
1	1	fx	air conditioning	not working
1	1	fx	speedometer	not working
1	1	fx	paint	fading
1	1	fx	mid-range sedan	disconnected
1	1		warning	light
1	1		map	light
1	1		suv	well maintain
1	1		aspects	satisfied
1	1		idea to rent a	excellent
1	1		drive	good
1	1		ford taurus	roomy
1	1		vacation	fun

To populate data pane:  
Make a selection in a table and click Display

# Data mining and text analytics

## Data Mining

- Use advanced analytical techniques on data
- Discover key relationships between variables
- Model effect of variables on outcomes
- Determine influence on outcomes
- Predict outcomes
- Apply models to new data in real-time

## Text Analytics

- Extract, analyse and create structure for unstructured data
- Integrate analysis results into operational systems
- Integrate analysis results into Business Intelligence applications
- Integrate analysis results with structured data and use as input for Data Mining
- Improves model accuracy



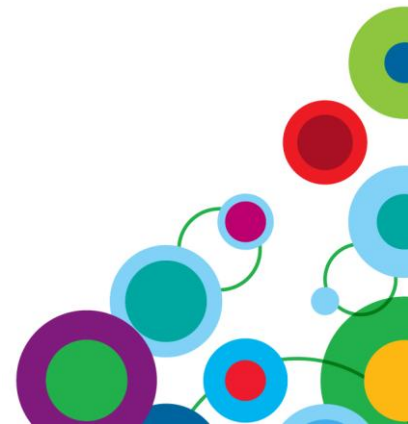


## IDC – Independent financial impact study



The median ROI for the projects that **incorporated predictive technologies was 145%**, compared with a median ROI of 89% for those projects that did not.

*Source IDC: “Predictive Analytics and ROI: Lessons from IDC’s Financial Impact Study”*





# Proven track record

- “**94% achieved a positive return on investment** with an average payback period of **10.7 months.**”
- “Returns were achieved through reduced costs, increased productivity, increased employee and customer satisfaction, and greater visibility.”
- “Flexibility, performance, and price were all key factors in purchase decisions.”

Nucleas Research, *An independent provider of Global Research and Advisory Services.*



“Reduced churn from 19 to 2%”



“35% reduction in mailing cost, 2X response rate, 29% more profit”



“100% increase in campaign effectiveness”



“30 Million Euro in new revenue”





## *Maximises revenue from targeted email marketing*

### Challenge

- What are the factors that are driving direct business through their e-commerce channels
- How to use wealth of customer data to tailor each marketing communication to a customer's unique needs

### Solution

- **Used IBM SPSS Modeler to develop customer profiles**
- **Used IBM SPSS Modeler to develop more accurate segmentation models**
- **Applied predictive intelligence to e-mail marketing campaign to target the right communication to the right customer**

### Results

- **Cost of e-mail marketing as a percentage of revenue (CPR) cut by 42% in 2009 vs. 2008**
- **Increased insight into customer activity drives loyalty**
- **Models and customer segmentation revealed where to target marketing spend**

**“The Customer Segmentation project allows us to keep in touch with our large database using cost-effective e-mail, but with all the benefits of a one-to-one relationship because we now have a clearly defined picture of each customer.”**

**↳ Chris Parker, direct analytics specialist at Avis Europe**





## *Gain and retain the right students*

### Challenge

- Access data held in multiple silos (admissions office, registrar, accounts receivable, etc.)
- Increase market visibility and target specific segments of prospective students

### Solution

- **Used IBM SPSS Modeler to access and consolidate multiple data stores to create a single view**
- **Created models for at-risk students, course placement, and student retention, and more**
- **Applied predictive intelligence across the student lifecycle**

### Results

- **In a declining business school market, saw 7.1% increased applications to business school**
- **21% annual increase in transfer students**
- **Decreased dropouts significantly by using predictive analytics to improve freshman placement**

**“These days, no meeting to make policy changes takes place without analysis based on predictive analytics”**

**↳ Jimmy Jung, Assistant VP for Enrollment Management**





## *Claims identification*

### Challenge

- Reduce payments on fraudulent claims
- Improve ability to collect payments from other insurance companies

### Solution

- **Used IBM SPSS Modeler to develop models of fraudulent claims**
- **Leveraged text analytics to interpret and analyse handwritten notes for use in investigation**
- **Extended use of predictive analytics beyond claims to customer retention and pricing analysis**

### Results

- **403% ROI in first 3 months**
- **Realised \$5 Million in benefit in the first year post-implementation**
- **Reduced cost of claims payment by enabling earlier, more targeted investigations**
- **Models deployed within call center to streamline claims process and gather the right data**

**“The relationship we have with our customers is put to the test when they file a claim, as they want a resolution so their lives can return to normal as quickly as possible. With SPSS, we can fast track valid claims or flag possible counterfeit claims for further review, saving our customers time and money.”**

**↳ Bill Dibble, SVP of Claims**





## *Increase public safety*

### Challenge

- Find innovative ways to fight escalating crime
- Find a cost-efficient way to analyse crime data, assess public safety risks, make intelligent decisions about personnel

### Solution

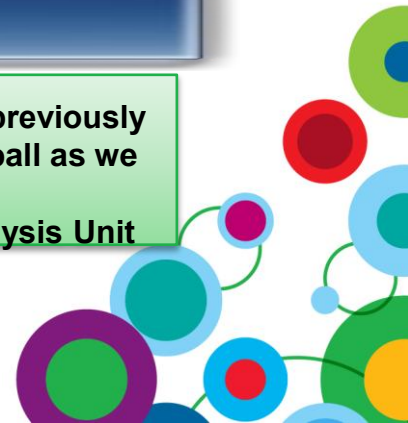
- **Analysts and officers users IBM SPSS Modeler to pore through data and find crime patterns and predict outcomes**
- **Forecast strategic positions for personnel and deployed “hot spot” maps to officers**
- **Used to identify key crime patterns to develop proactive policing strategies**

### Results

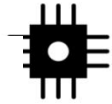
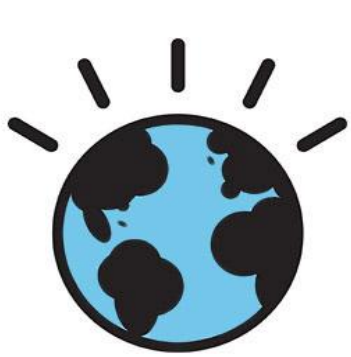
- **Dramatic reduction in crime between 2006 and 2007 despite economic conditions**
- **New Year’s Eve test saw 246% increase in weapon seizure, 49% decrease in gunfire, and \$15,000 savings in overtime**
- **Gives even rookie officers veteran-like insight into crime data**

**“IBM SPSS Modeler and data mining represent a revolution in our ability to access previously unobtainable data, and pull meaning and value from it. This is as close to a crystal ball as we are ever going to get.”**

**→ Colleen McCue, program manager for the Department’s Crime Analysis Unit**



# Drive data in decision making



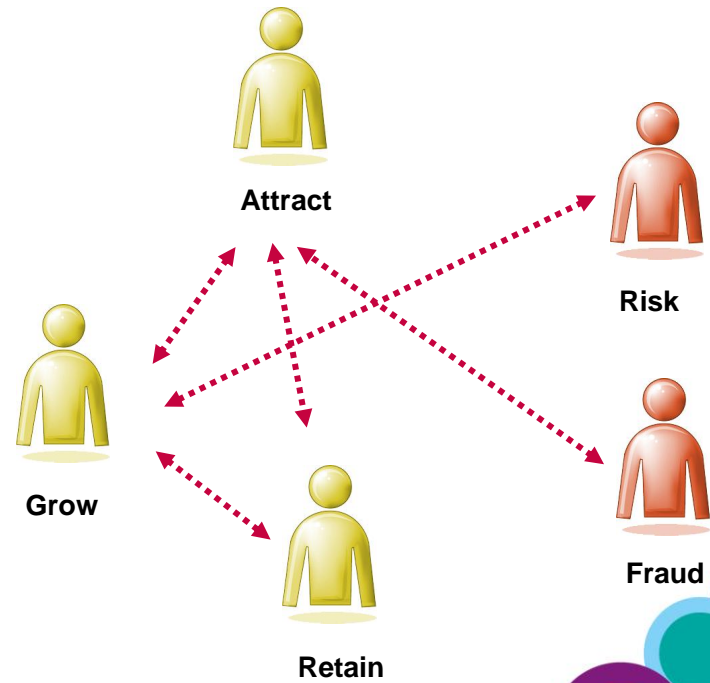
Instrumented



Interconnected



Intelligent



# **WIN** *an Apple<sup>®</sup> iPad*

Please remember to complete your session evaluation online at the Communication Station or point your Smart Phone browser to:

**[www.spss.com/goldcoast](http://www.spss.com/goldcoast)**



For a chance to win an Apple<sup>®</sup> iPad

