

# WEBSITE AND EMAIL PERSONALIZATION

## THE NEXT WAVE

Informs Presentation  
January 13, 2010

**Data Mining Technologies Inc.**  
*Knowledge Discovery From Databases*

Michael Gilman, Ph.D.  
Copyright Data Mining Technologies Inc 2010  
631 -692-4400 ext. 100  
[www.data-mine.com](http://www.data-mine.com)



# MICHAEL GILMAN

- CEO of Data Mining Technologies Inc.
- Corporate management, data mining, business analytics, project management, systems analysis, and university teaching.
- PhD Management Science, Applied Mathematics NYU
- MS Computer Science NYU
- BEE CUNY Electrical Engineering

# DISCOVERING THE FUTURE TODAY

## WITH *NUGGETS*<sup>™</sup> BY DATA MINING TECHNOLOGIES

 CHASE The right relationship is everything.®

NATIONAL RECOGNITION PRODUCTS



POST TIME TECHNOLOGIES

 March of Dimes  
*Saving babies, together*

 **asL**  
AMERICAN STUDENT LIST

 **ARGI**  
AUTOMATED RESOURCES GROUP, INC.

*info*USA  
The Premier Source of Sales & Marketing Solutions

The **ROBINSON** Company

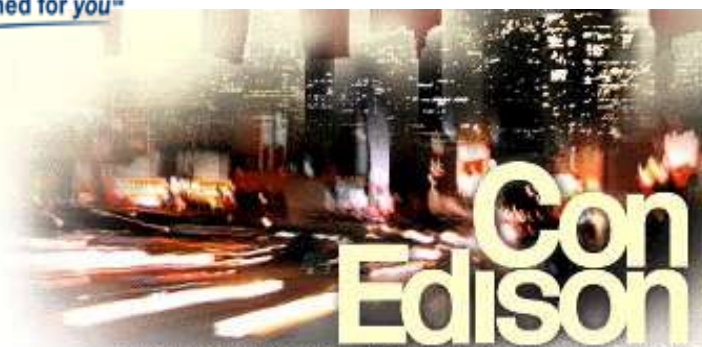
**IBHDS** BLUE HILL DATA SERVICES, INC.  
*No Compromise Solutions*

Celebrity **X** Cruises®  
*Designed for you™*

**ASCEND ONE**

**Musella  
Medical  
Research**

 **Polytechnic**  
UNIVERSITY



THE COMPANY YOU KNOW. THE PEOPLE YOU TRUST.  
A Subsidiary of Consolidated Edison, Inc.

# WHAT WE'LL COVER

---

- ❖ Personalization – The coming 3<sup>rd</sup> wave
- ❖ Methodologies – Things you should know about how to achieve effective personalization
- ❖ 10 Challenges of Predictive Modeling



# PERSONALIZATION

---

- ❖ What is it
- ❖ How widespread
- ❖ Why is it used - benefits
- ❖ Where it is used
- ❖ How is it used
- ❖ Technical challenges



# PERSONALIZATION

---

- ❖ **What is it**
- ❖ How widespread
- ❖ Why is it used - benefits
- ❖ Where it is used
- ❖ How is it used
- ❖ Technical challenges



# WHAT IS PERSONALIZATION?

---

**Personalization** involves using technology to accommodate the differences between individuals. Once confined mainly to the Web, it is increasingly becoming a factor in education, healthcare (i.e. personalized medicine), and both "business to business" and "business to consumer" settings. (Wikipedia)



# PERSONALIZATION DATA

---

**Personal** - recommend things based on the individual's past behavior (transactional)

**Group**- recommend things based on the past behavior of similar individuals (demographic/psychographic)

**Item** - recommend things based on the item itself

★ A combination of the approaches above





# PERSONALIZATION

---

- ❖ What is it
- ❖ **How widespread**
- ❖ Why is it used - benefits
- ❖ Where it is used
- ❖ How is it used
- ❖ Technical challenges



# HOW WIDESPREAD?

---

Personalization has shown slow but steady progress in 2009. We consider personalization to be a key facet of the evolving Web.



*Source: ReadWriteWeb's Top 5 Web Trends of 2009:*



# HOW WIDESPREAD?

---

- ❖ e-commerce - low
- ❖ email - low
- ❖ Healthcare - low
- ❖ Institutional – low



*My take: still in infancy*



# AMAZON

---

## Collaborative Filtering

### Sample purchase statistics

Customer	Item 1	Item 2	Item 3
John	Bought it	Didn't buy it	Bought it
Mark	Didn't buy it	Bought it	Bought it
Lucy	Didn't buy it	Bought it	Didn't buy it



# AMAZON

[Share with Friends](#)

## frequently Bought Together



Price For All Three: **\$75.05**

[Show availability and shipping details](#)

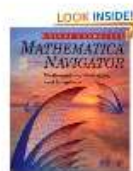
- This item:** Mathematica DeMYSTiFieD by Jim Hoste
- [Schaum's Outline of Mathematica, 2ed \(Schaum's Outline Series\)](#) by Eugene Don
- [Mathematica Navigator: Mathematics, Statistics and Graphics, Third Edition](#) by Heikki Ruskeepää

## customers Who Bought This Item Also Bought

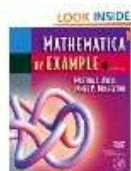
Page



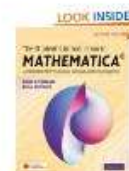
[Schaum's Outline of Mathematica, 2ed \(Schaum's Outline Series\)](#) by Eugene Don  
★★★★☆ (14) \$12.89



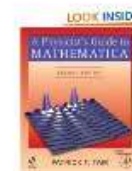
[Mathematica Navigator: Mathematics, Statistics and Graphics, Third Edition](#) by Heikki Ruskeepää  
★★★★★ (15) \$45.69



[Mathematica by Example, Fourth Edition](#) by Martha L. Abell  
★★★★☆ (4) \$40.92



[The Student's Introduction to MATHEMATICA... \(The Student's Introduction to Mathematica\)](#) by Bruce F. Torrence  
★★★★★ (4) \$45.00



[A Physicist's Guide to Mathematica, Second Edition](#) by Patrick Tam  
★★★★★ (3) \$53.47

# PERSONALIZATION

---

- ❖ What is it
- ❖ How widespread
- ❖ Why is it used - benefits**
- ❖ Where it is used
- ❖ How is it used
- ❖ Technical challenges



# PERSONALIZATION - WHY?

---

Increase Profitability



# PERSONALIZATION - WHY?

---

Increase Receptivity





# PERSONALIZATION - WHERE?

---

## Target

### Applications

❖ e-commerce

❖ email

❖ Healthcare

❖ Institutional



# PERSONALIZATION - WHERE?

---

❖ **ecommerce**

❖ **email**

❖ **Healthcare**

❖ **Institutional**



# THE BIG IDEA IN WEB PERSONALIZATION

How Do You Make Money in ecommerce

Get them to the site

Adwords, etc

Organic Search

Get them to buy



**Personalize**



# ANATOMY OF A SITE VISIT

- Search or Enter URL of site
  - From Ad \$\$
  - Directly
- Get to Landing page directly or from link or from search box on site home page
- Convert or Out
  - View standard page – low conversion rate
  - View personalized page – higher conversion rate
  - Possible negative profit from visit



# PERSONALIZATION - WEBSITE

---

## Personalized Landing Pages

 ecommerce

Creative choice1 .“Price Message”

Creative choice2. “Full Inventory Message”

Creative choice3. “Sales Message”

Creative choice4. “Quality of Service Message”



# PERSONALIZATION - WEBSITE

---

## ► ecommerce

### Personalized Landing Pages

#### Creative Page Types

“Price Message”

“Full Inventory Message”

“Sales Message”

“Quality of Service Message”

#### Landing Page Components

Headline

Body Copy

Graphic



# PERSONALIZATION - WEBSITE

## eCommerce

### Landing Page Components

Headline

Body Copy

Graphic

The screenshot displays a product page for a TaylorMade r7 Iron Set. The navigation bar includes a search bar, 'GOLF STORE', and various category links like 'PRE-OWNED', 'CLUBMAKING', 'WOMEN', and 'TENNIS'. A large red banner at the top of the main content area reads 'WINTER SALE & CLEARANCE EVENT - Save up to 60% on all the best brands'. Below this, the breadcrumb trail shows the user's path: 'Golf Store Text Search > 10SALCLEAR (remove) \*10SALCLEAR\* > Clubs (remove) > TaylorMade r7 Iron Set 4-PW, GW with Steel Shafts'. The product image shows a close-up of the iron head. To the right, the product title is 'TaylorMade r7 Iron Set 4-PW, GW with Steel Shafts', followed by a 4.9 out of 5 star rating and a 'Write a review' link. The price is listed as '\$399.99 Was: \$699.99 Save: \$300.00'. Below the price, there are dropdown menus for 'HAND' (Right) and 'FLEX' (Flex), along with 'Buy With Confidence' and 'Special Financing' options. A 'QUANTITY' selector is set to '1', and the 'SUB TOTAL' is '\$399.99'. An 'ADD TO CART' button is prominently displayed in a red box. At the bottom, there is a note '† Excludes Taxes and Shipping Costs' and a 'PICK IT UP AT THE STORE' option with a zip code field.

# PERSONALIZATION - WEBSITE

ecommerce

Landing Page Components

Headline

Body Copy

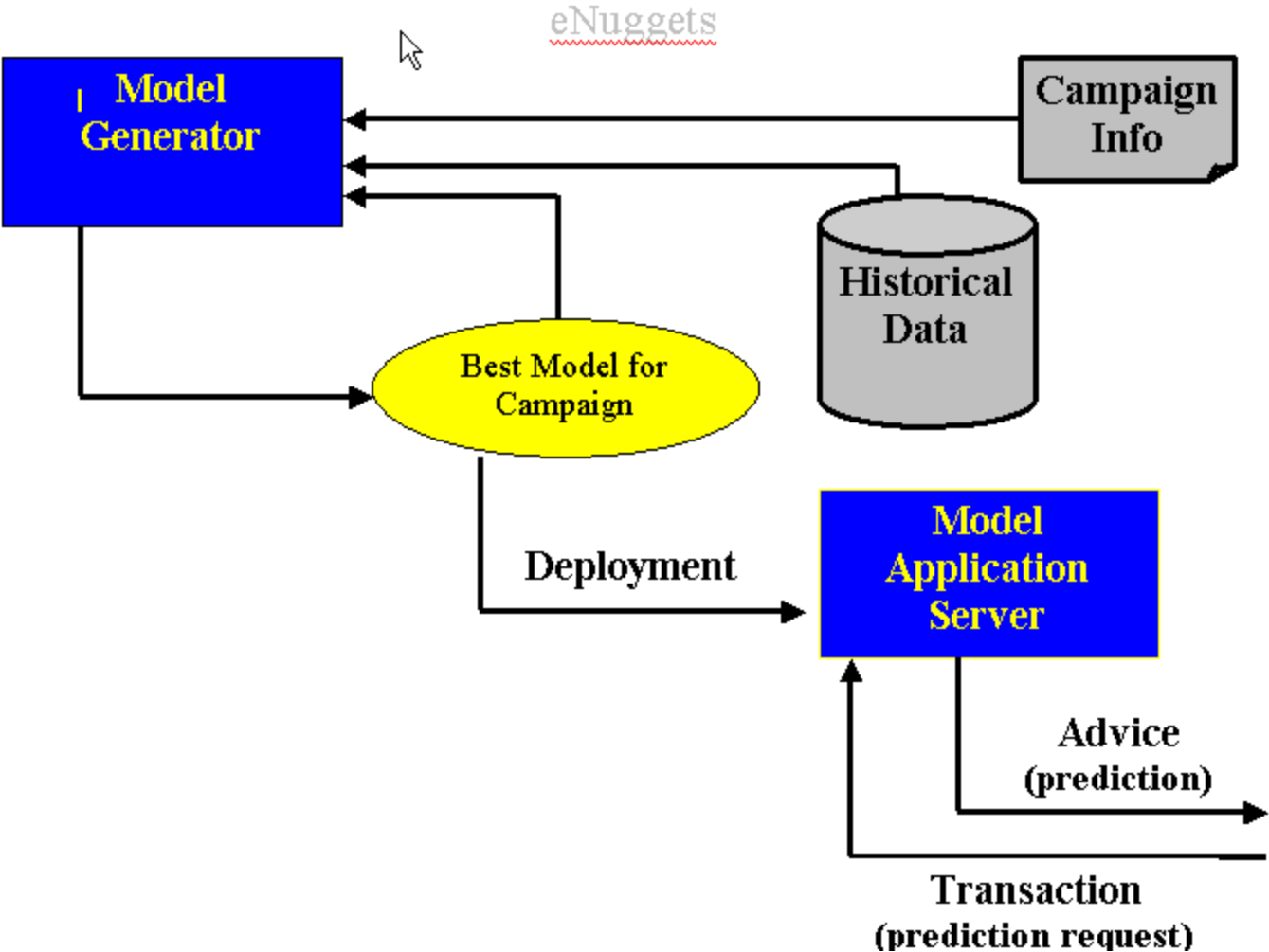
Graphic

The screenshot shows a web browser window displaying the 'GOLF STORE' website. At the top, there is a search bar and navigation links for 'GOLF STORE', 'SALE & CLEARANCE', 'PRE-OWNED', 'CLUBMAKING', 'WOMEN', 'LOGO', and 'TENNIS'. Below this is a secondary navigation menu with categories like 'Golf Clubs', 'Shoes', 'Apparel', etc. A prominent red banner reads 'WINTER SALE & CLEARANCE EVENT - Save up to 60% on all the best brands'. The main content area shows a search path: 'Golf Store Text Search > 10SALCLEAR (remove) \*10SALCLEAR\* > Clubs (remove) > TaylorMade r7 Iron Set 4-PW, GW with Steel Shafts'. The product title is 'TaylorMade r7 Iron Set 4-PW, GW with Steel Shafts', with a rating of 4.9 out of 5. The price is '\$399.99 Was: \$699.99 Save: \$300.00'. The product options section includes 'HAND' (Right), 'FLEX' (Flex), and 'QUANTITY' (1). A large image of a woman in a red bikini is positioned to the left of the product details. The 'ADD TO CART' button is highlighted in red.

New Graphic



# eNuggets Data Flow



# PERSONALIZATION - WHERE?

---

❖ e-commerce

❖ email

❖ Healthcare

❖ Institutional



# PERSONALIZATION - EMAILS

---

## Direct Mail

Issue: Who do we mail to?

## Email Mail

What do we say?

How often do we say it?



# PERSONALIZATION - EMAILS

Use one of several possible creative messages selected by a model

## Creative Themes

- “Price Message”
- “Full Inventory Message”
- “Sales Message”
- “Quality of Service”



# PERSONALIZATION - EMAILS

---

Makeup of creative for email ad

Headline

Body Copy

Offer

Call to Action



# PERSONALIZATION - WHERE?

---

## email

### Choice of creative message

1. Sale on women's sweaters
2. Store wide Labor Day sale

### Choice of product offering

#### Debt Remediation Plans

- Package 1
- Package 2
- Package 3

#### Extended Warranty for Cars

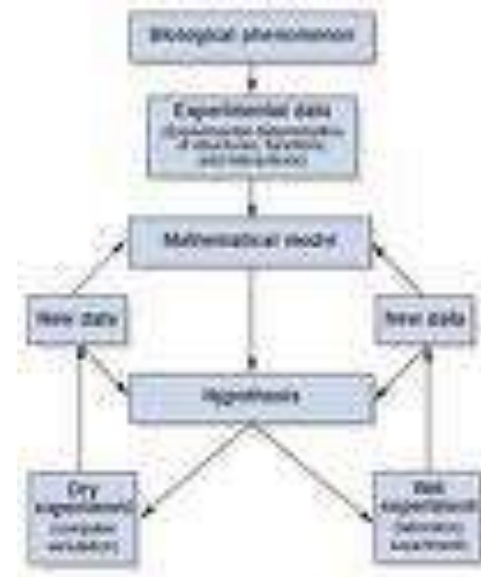
- Package 1
- Package 2
- Package 3
- Show all three
- Default

E



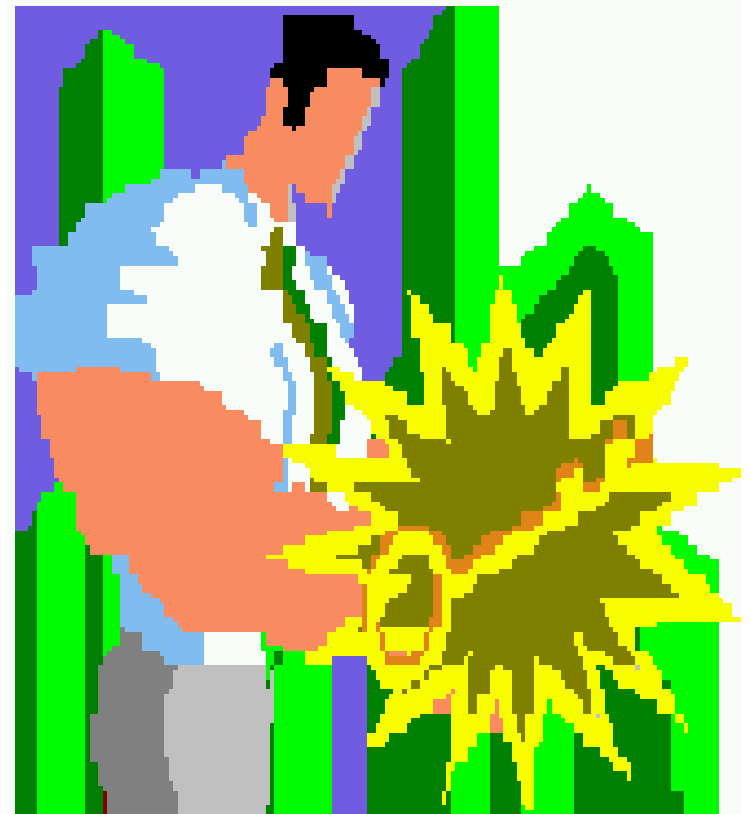
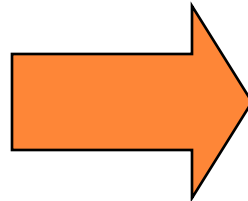
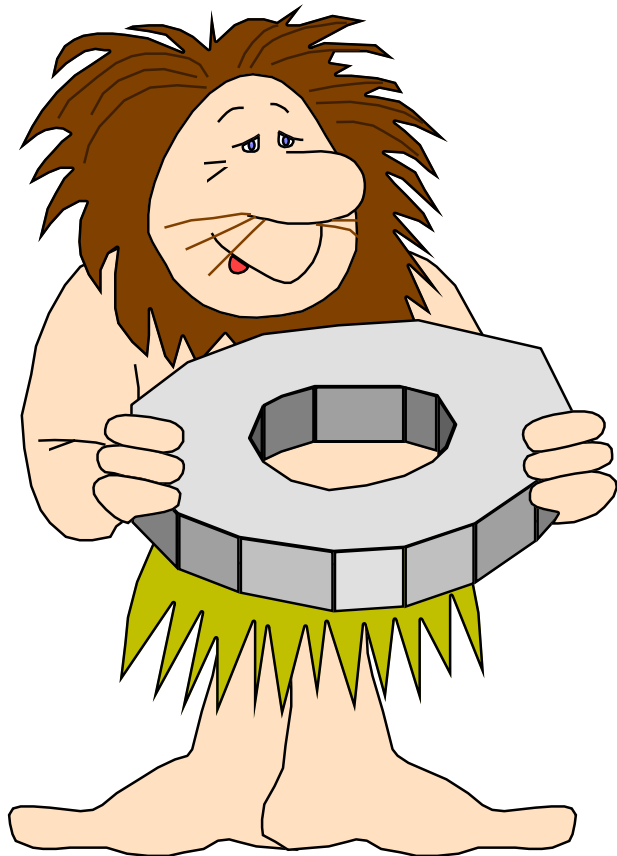
# POPULAR MODELING METHODS

- Statistics
- Neural Networks
- Tree Builders
- **Rule Induction**



# NUGGETS®

## PERSONALIZATION- THE NEXT STEP...





# CHALLENGE OF DATA

1. Highly non-linear
2. Predictor attributes are often correlated
3. Data is mixture of numeric (cardinal, ordinal), nominal, longitudinal
4. Lurking variables
5. Often non-stationary
6. Sparse
7. Noisy
8. Patterns are many – large and small
9. Methods find correlations not usually causes
10. Correlation measures defined for linear relationships

Most methods don't deal with above easily



# SUPPLEMENTARY SLIDES

## DATA MINING OVERVIEW



# WHAT IS DATA MINING?

---

An information extraction activity which has as its goal the discovery of hidden facts contained in databases. It finds patterns and subtle relationships in data, inferring rules and generalizations that allow the prediction of future results. To be a true *knowledge discovery* method, a data mining tool should unearth information *automatically*.



# WHAT DATA MINING IS NOT

- OLAP
- Other Querying



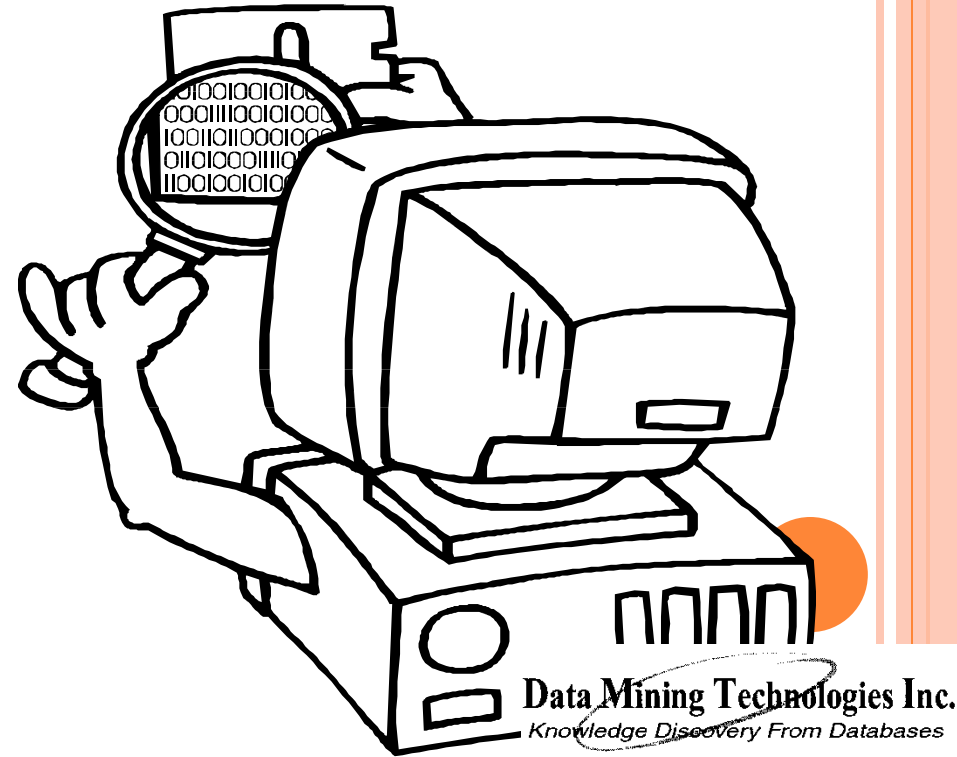
# *WITH QUERYING/OLAP PROGRAMS*

You formulate specific questions

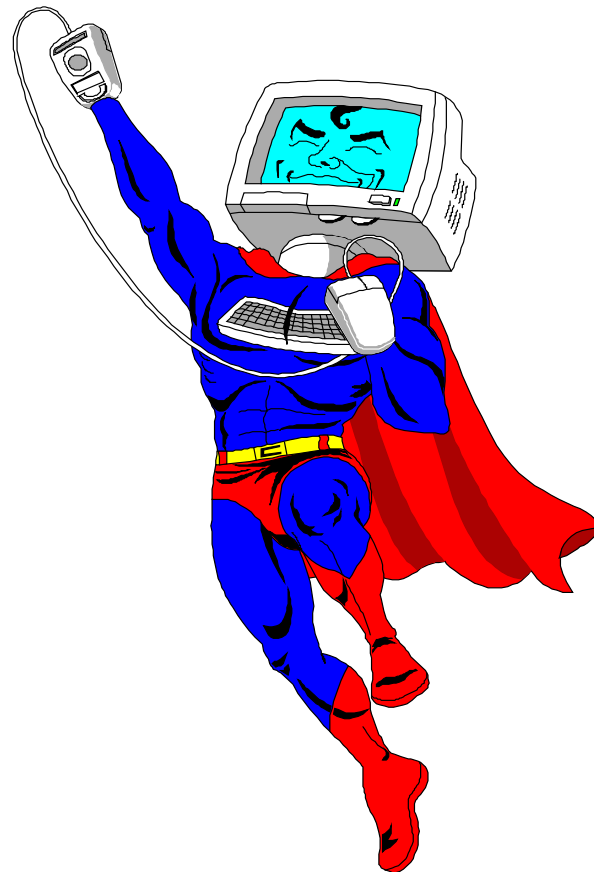


# BUT WITH DATA MINING

- You Formulate General Questions



# AN EXAMPLE



# QUESTION:

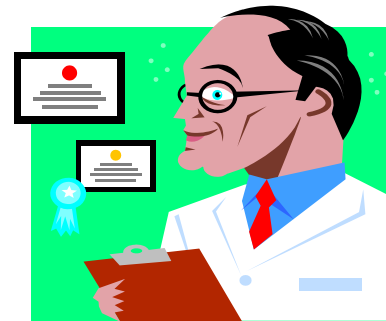


How can we keep healthcare quality high and keep costs down ?





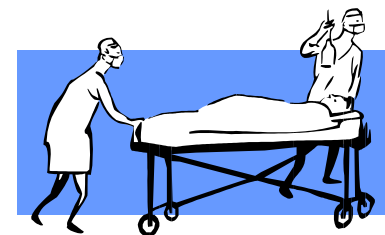
INPUT DATA:  
FILE CONTAINING CLINICAL DATA AND  
COSTS



# RESULTS

- Model containing rules showing what is best of breed treatment for each case and why

*If diagnosis = Congestive HF and Age =60-70 and previous. bypass = yes and . . . Then BOB Treatment = aortic stent*



# MODEL TYPES

- Response
- Cloning
- Attrition
- Cross Sell
- Upsell
- High Lifetime value
- Matrix

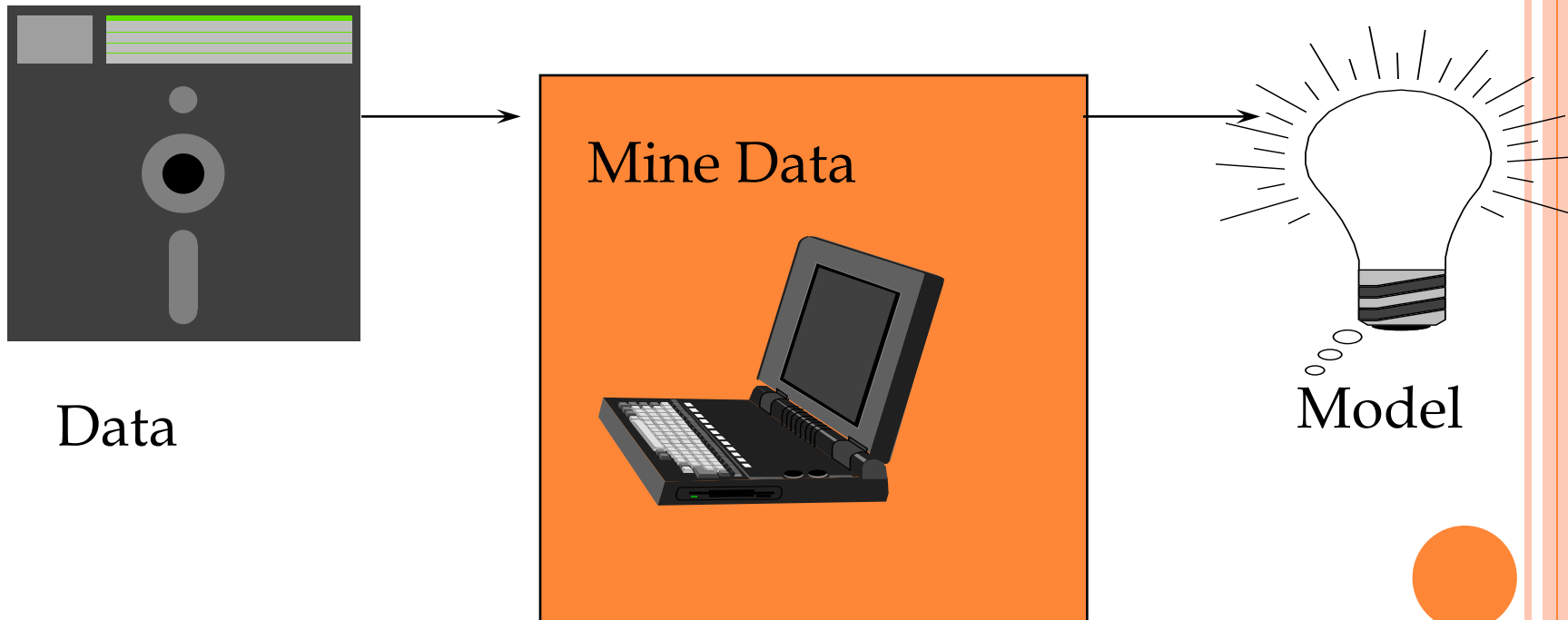


# STEPS IN DATA MINING

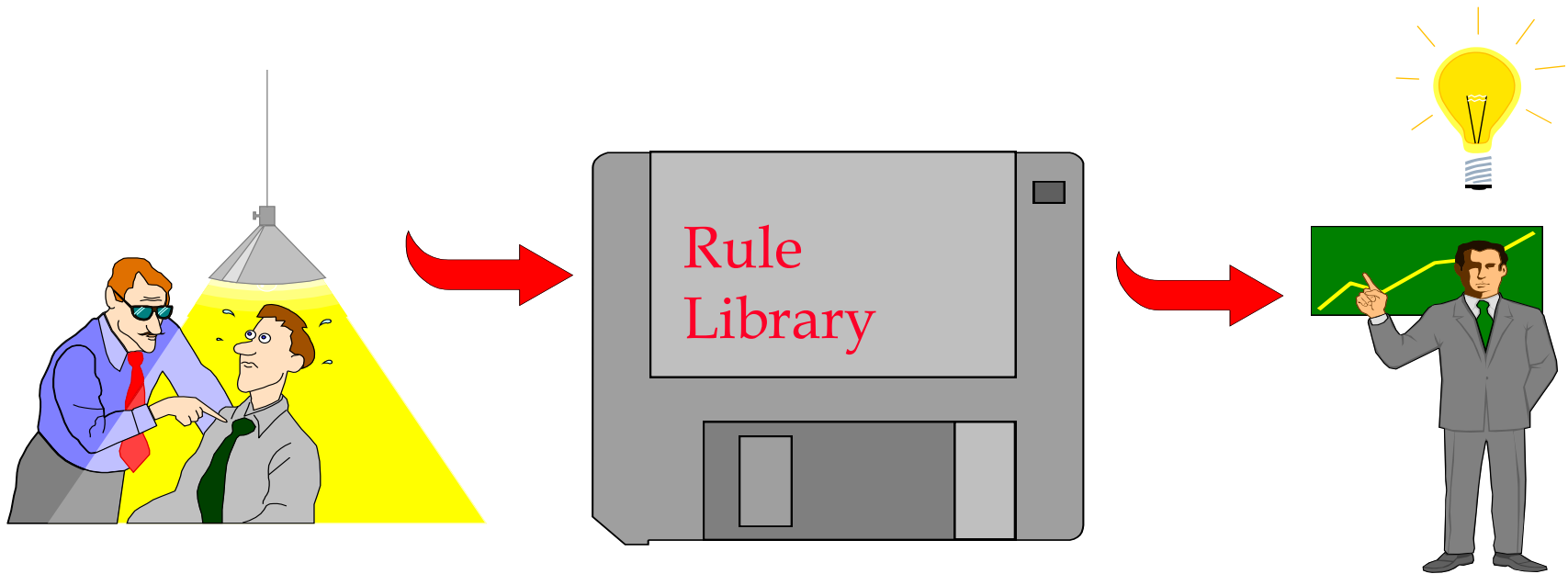
- Define the problem goals
- Identify data sources



# THEN BUILD THE MODEL

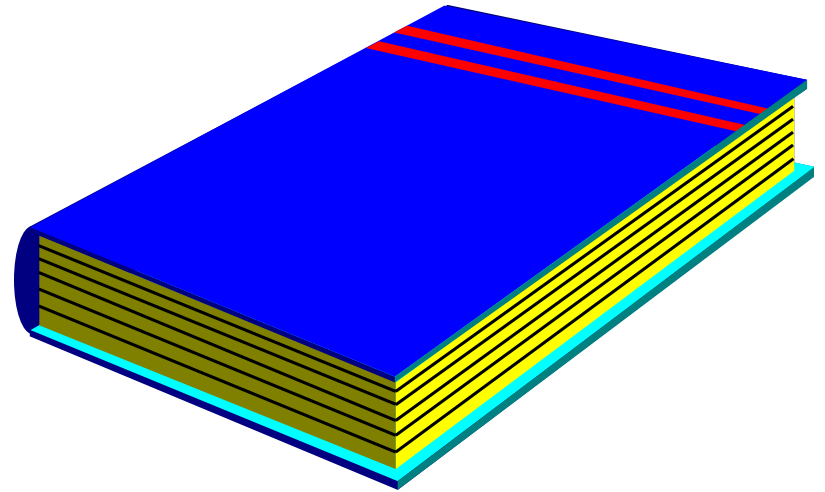


# THE MODEL NOW REPRESENTS THE HIDDEN PATTERNS. . .



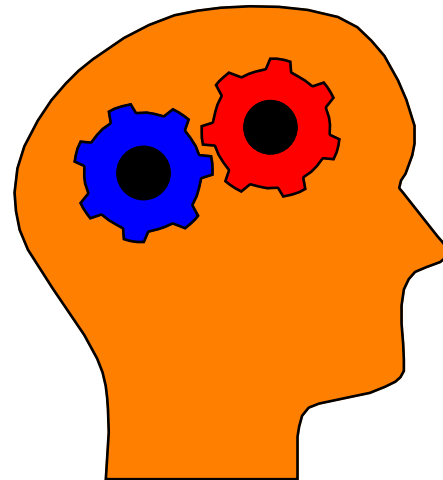
## GENERATE RESULTS

- New data are then predicted using the model
- or Generalizations are made



# VALIDATE THE RESULTS

- Holdout
- Score
- Tune

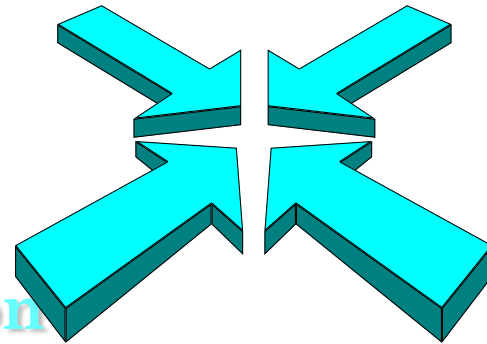




# PRINCIPAL DATA MINING TECHNIQUES

- Neural Nets
- Statistics
- Tree Builders

○ **Nuggets** Rule Induction



# METHOD PROS/CONS

- Neural Networks

- Pros - Aimed at numerical problems
- Cons - Hard to use, black box, art vs. science, overfitting issue



# METHOD PROS/CONS

## ○ Statistics

- Pros - Established technology
- Cons - Needs assumptions, nominal variable handling, tricky to use, smoothes out patterns



# METHOD PROS/CONS

- Tree Builders (Cart, Chaid, etc.)
  - Pros – Gives rules, not black box
  - Cons – Not very accurate, small number of variables, “greedy algorithm”



# METHOD PROS/CONS

- True Rule Induction - **Nuggets**
  - Pros - Understandable results, easy to use, granular, highly accurate, nominal variables
  - Cons - Need to bin numerical data



# Nuggets



*Nuggets* uses proprietary search algorithms to intelligently prospect data for valid hypotheses.

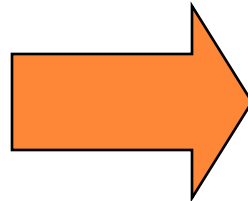
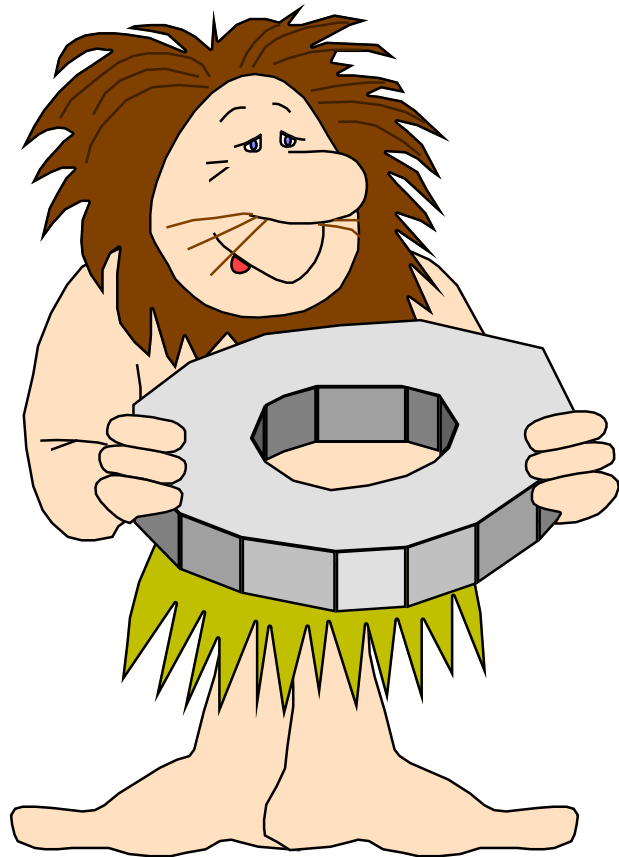
In the act of searching, the algorithms “learn” about the training data as they proceed.

The result is a very fast and efficient discovery strategy that does not preclude any potential rule or generalization from being found.



# NUGGETS®

## RULE INDUCTION - THE NEXT STEP...



## **NUGGETS**<sup>TM</sup> ITS UNIQUE

---

❖ Power to find the patterns from data that other methods miss

❖ Automatic rule generation in understandable English “if-then” rules

❖ Ability to handle complex non-linear relationships

❖ Handles both missing and noisy data

❖ Assists in finding data errors





## ADVANTAGES CONTINUED...

- ✦ fast
- ✦ handles up to 50,000 variables
- ✦ easy to use
- ✦ highly accurate
- ✦ no statistics
- ✦ not black box
- ✦ provides a validation module, other tools
- ✦ reverse engineers information implicit in databases
- ✦ desk top or real time



# DATA MINING

- The future of information technology in business and science



# WEBSITE AND EMAIL PERSONALIZATION

## THE NEXT WAVE

Informs Presentation  
January 13, 2010

**Data Mining Technologies Inc.**  
*Knowledge Discovery From Databases*

Michael Gilman, Ph.D.  
Copyright Data Mining Technologies Inc 2010  
631 -692-4400 ext. 100  
[www.data-mine.com](http://www.data-mine.com)



# DISCOVERING THE FUTURE TODAY

## WITH **NUGGETS**<sup>™</sup> BY DATA MINING TECHNOLOGIES

**ASCENDONE**

NATIONAL RECOGNITION PRODUCTS



**CHASE** The right relationship is everything.®



POST TIME TECHNOLOGIES



**infoUSA**  
The Premier Source of Sales & Marketing Solutions

The **ROBINSON** Company

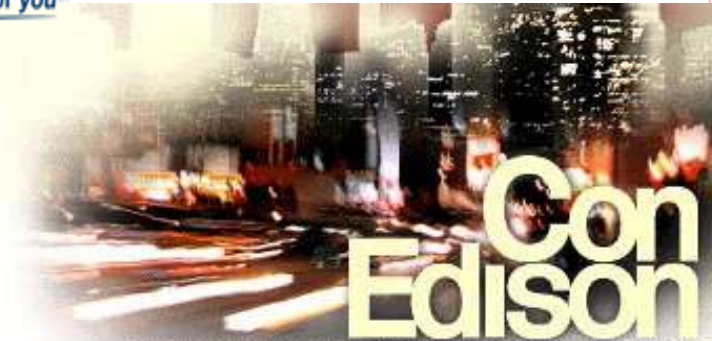
Celebrity **X** Cruises®  
Designed for you™



**BLUE HILL DATA SERVICES, INC.**  
*No Compromise Solutions*

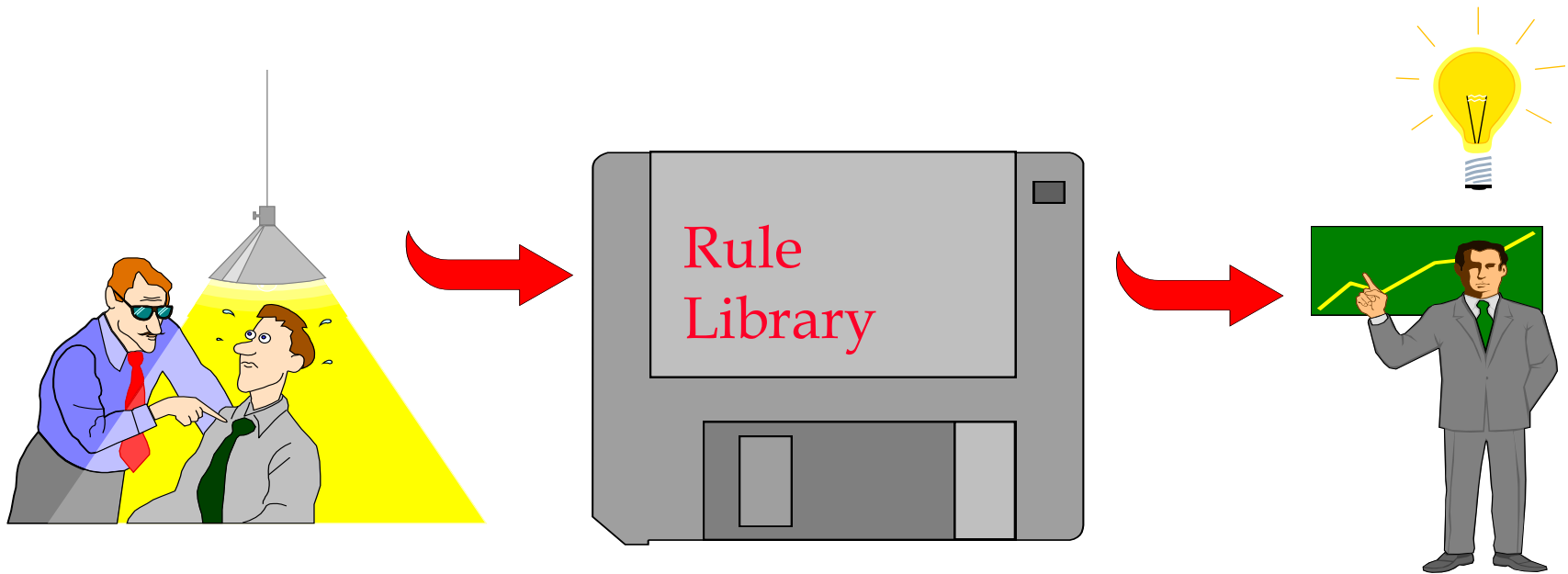


**Musella  
Medical  
Research**



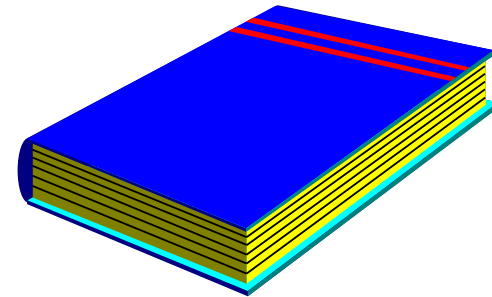
THE COMPANY YOU KNOW. THE PEOPLE YOU TRUST.  
A Subsidiary of Consolidated Edison, Inc.

# THE MODEL NOW REPRESENTS THE HIDDEN PATTERNS. . .



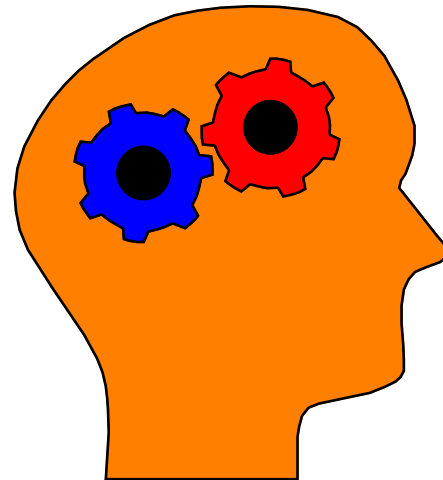
## GENERATE RESULTS

- New data are then predicted using the model
- or Generalizations are made



# VALIDATE THE RESULTS

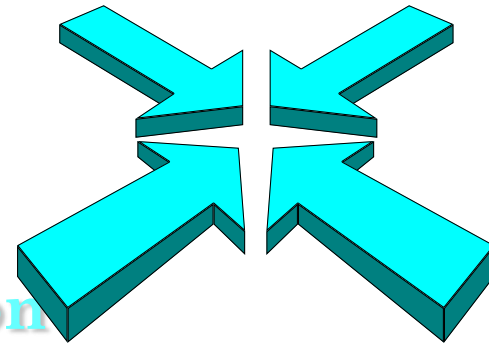
- Holdout
- Score
- Tune



# PRINCIPAL DATA MINING TECHNIQUES

- Neural Nets
- Statistics
- Tree Builders

○ **Nuggets** Rule Induction





# METHOD PROS/CONS

## ○ Neural Networks

- Pros - Aimed at numerical problems
- Cons - Hard to use, black box, art vs. science, over fitting issue



# METHOD PROS/CONS

## ○ Statistics

- Pros - Established technology
- Cons - Needs assumptions, nominal variable handling, tricky to use, smoothes out patterns



# METHOD PROS/CONS

- Tree Builders (Cart, Chaid, etc.)
  - Pros – Gives rules, not black box
  - Cons – Not very accurate, small number of variables, “greedy algorithm”



# METHOD PROS/CONS

- True Rule Induction - **Nuggets**
  - Pros - Understandable results, easy to use, granular, highly accurate, nominal variables
  - Cons - Need to bin numerical data





**Nuggets**

*Nuggets* uses proprietary search algorithms to intelligently prospect data for valid hypotheses.

In the act of searching, the algorithms “learn” about the training data as they proceed.

The result is a very fast and efficient discovery strategy that does not preclude any potential rule or generalization from being found.

## **NUGGETS**<sup>TM</sup> ITS UNIQUE

---

❖ Power to find the patterns from data that other methods miss

❖ Automatic rule generation in understandable English “if-then” rules

❖ Ability to handle complex non-linear relationships

❖ Handles both missing and noisy data

❖ Assists in finding data errors

## ADVANTAGES CONTINUED...

- ✘ fast
- ✘ handles up to 50,000 variables
- ✘ easy to use
- ✘ highly accurate
- ✘ no statistics
- ✘ not black box
- ✘ provides a validation module, other tools
- ✘ reverse engineers information implicit in databases
- ✘ desk top or real time

# DISCUSSION/QUESTIONS

