Data Mining and its Application in Marketing and Business

Credit: Gaby Matalon

What is Data Mining?

- The process of analyzing data from different perspectives and summarizing it into useful information
- It uncovers patterns in a large set of data
- Growing industry to target their products and advertisements towards consumers based on data mining

Why Data Mining?

- Almost everything we do leaves electronic data behind

- Needs to extract useful information from data in order to interpret the data

- Data mining helps speed up the process of finding relationships and patterns in raw data

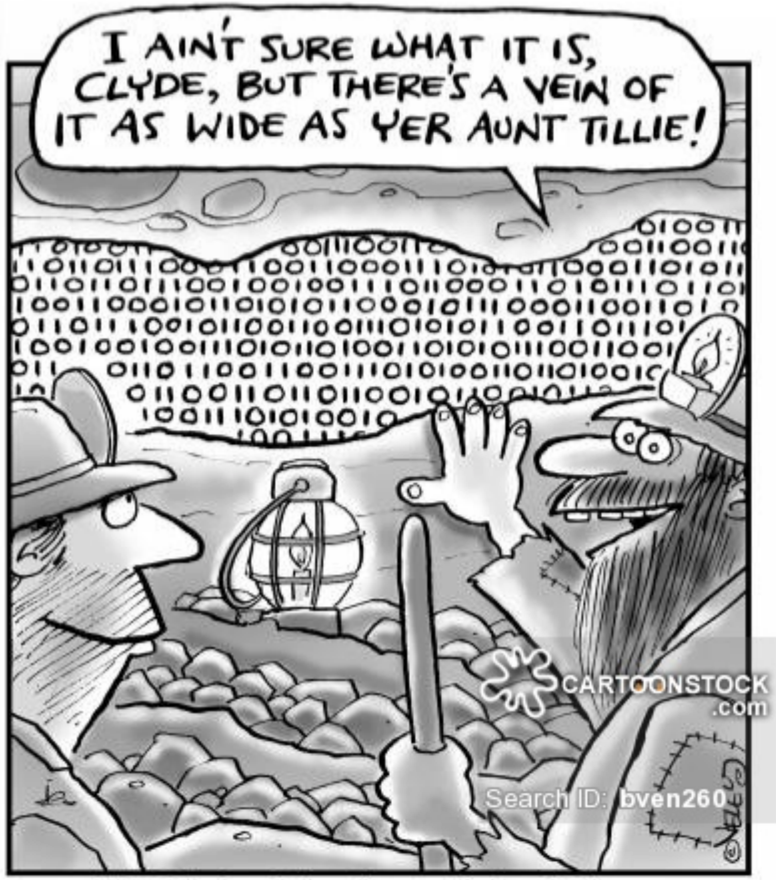
Uses of Data Mining

- Healthcare
- Finance
- Retail & E-Commerce
  - Learn about consumer preferences
- Countless others!

OUR CONSULTANT HAS BEEN MINING DATA ALL DAY.
THE RESULTS ARE QUITE SHOCKING.
www.dilbert.com
scottadams@aol.com
ACCORDING TO THE DATA, SALES ARE ALWAYS HIGHEST WHEN I DO THIS...
4/5/00 © 1999 United Feature Syndicate, Inc.

History of Data Mining

- The term “data mining” is relatively new but the concepts have been around for many years
- Classical statistics, artificial intelligence and machine learning culminated over the years and evolved into data mining


I AIN'T SURE WHAT IT IS, CLYDE, BUT THERE'S A VEIN OF IT AS WIDE AS YER AUNT TILLIE!
CARTOONSTOCK .com
Search ID: bven260

The birth of the data mining industry (circa 1880)

History of Data Mining (Cont.)

- Data Collection (1960s)- process of storing information on computers
  - Technology- computers, tapes and disks

- Data Access (1980s)-the introduction of structured query languages and relational databases helped us learn more about data
  - Data available at record level dynamically

History of Data Mining (Cont.)

- Data Warehousing and Decision Report (1990s)-the process of centralized data management and retrieval
  - Maintaining a central location for all organizational data
  - Helps you analyze data and concentrate on very specific characteristics
  - Dynamic data delivery at multiple levels

- Data Mining (present)- generalizing patterns, predictive

Influential People/Events

- In 1975, John Henry Holland wrote *Adaptation in Natural and Artificial Systems*, a book on genetic algorithms – start in data mining
- 1990s, the term “data mining” appeared in the database community for the first time
- In 2001, William S. Cleveland introduced data mining as an independent discipline
- DJ Patil became the first Chief Data Scientist in the White House in February 2015

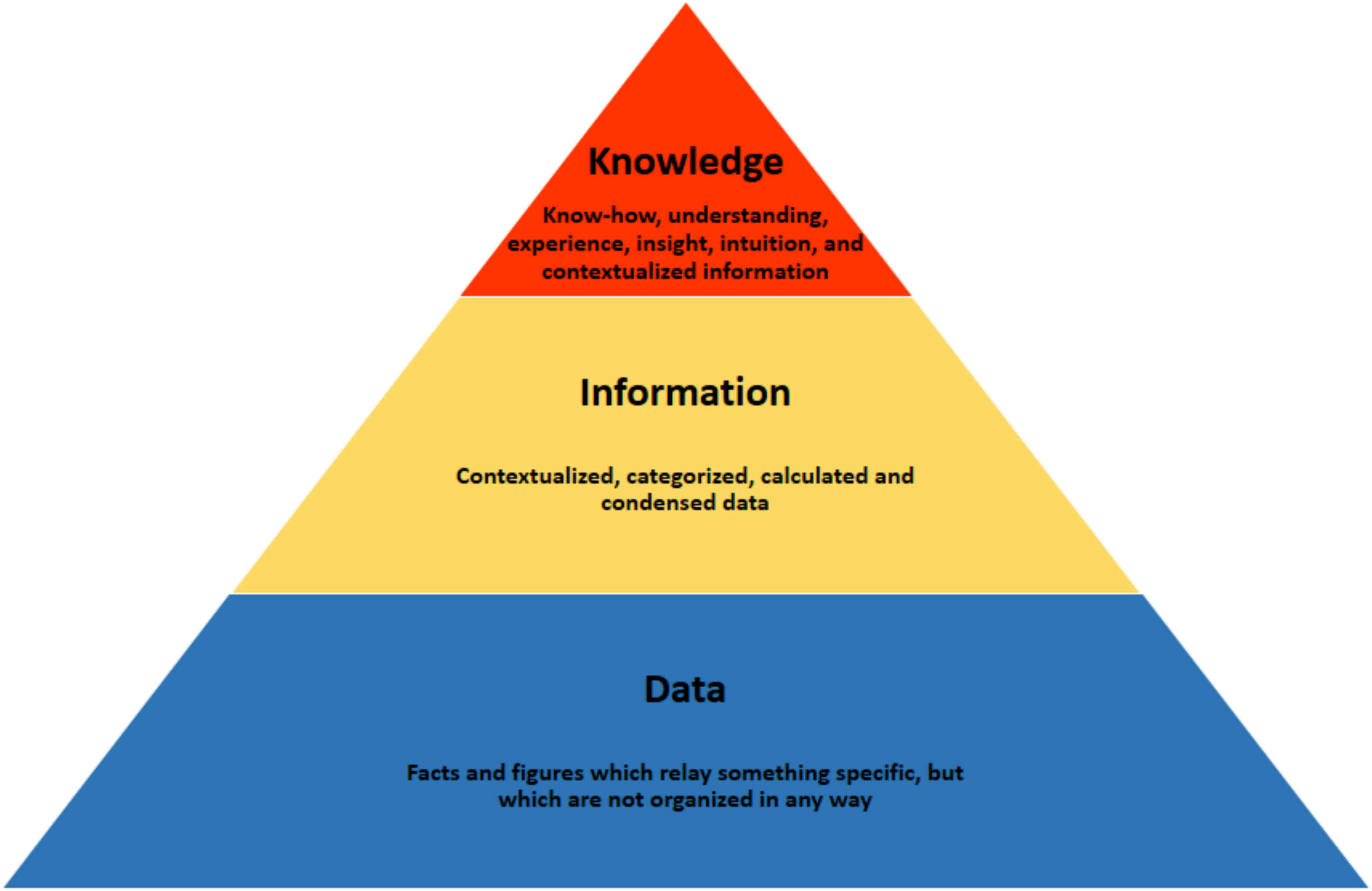
Terminology

- Data: facts, numbers or text that can be processed by a computer

- Information: the patterns, associations and relationships of data

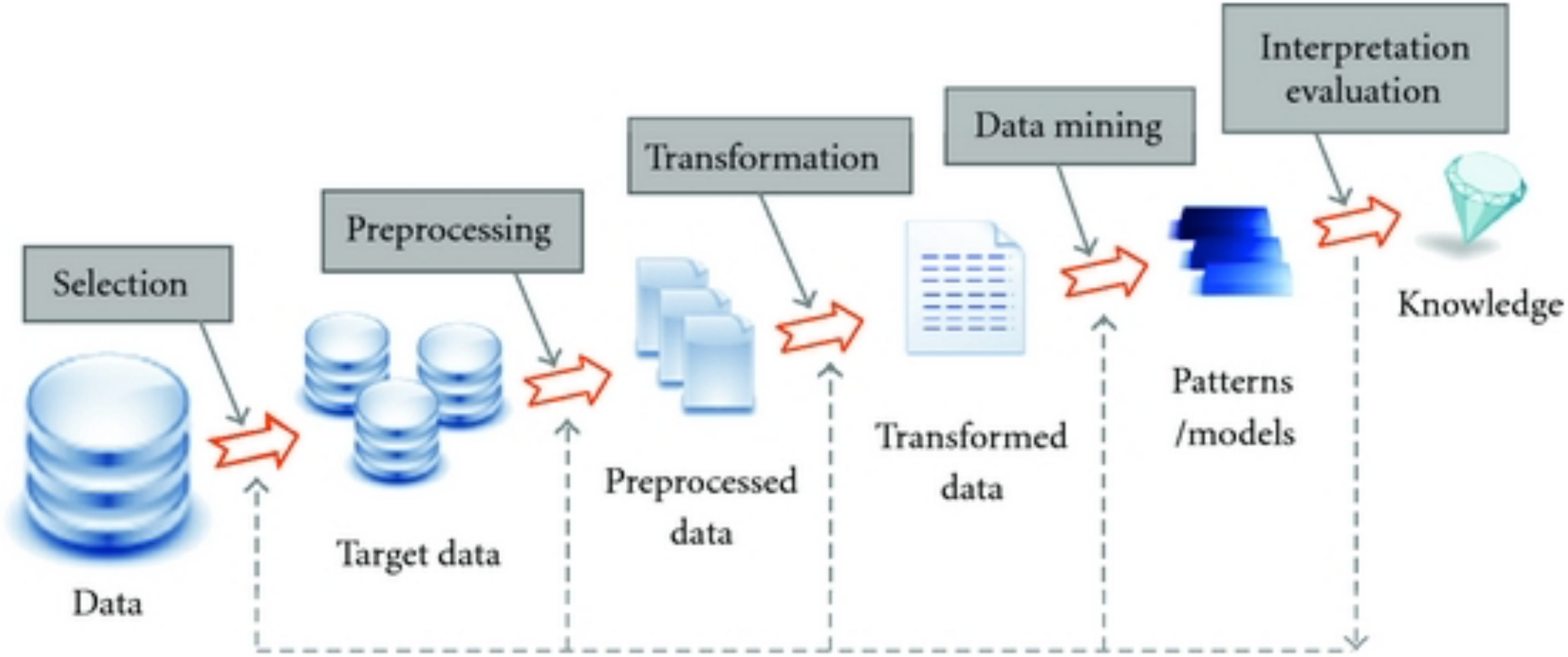
- Knowledge: understanding of a subject, synthesise information to gain knowledge about historical patterns and future trends

Data, Information, vs. Knowledge



Knowledge
Know-how, understanding, experience, insight, intuition, and contextualized information
Information
Contextualized, categorized, calculated and condensed data
Data
Facts and figures which relay something specific, but which are not organized in any way

Data Mining in the Knowledge Discovery Process



How Data Mining Works

- Need a clear understanding of data in order to be able to use the information effectively

- To do this, classify data into these groups:
  1. *Classes*
  2. *Clusters*
  3. *Associations*
  4. *Sequential Patterns*

Classes

- Classes are groups where the data shares characteristics
    - Example- A class for a company like Netflix would be the customers who all watched a certain movie

Clusters

- Very similar to classes but with additional attributes (ie. logical relationships & consumer preferences)
    - Example- product recommendations, people who bought this product also bought…

- In terms of business this can be the most helpful/effective way to classify data

Associations

- Takes clustering even further
- Use data to find relationships/patterns that would often go unnoticed
    - Example: finding a connection between buying two unrelated products
- Associations can help businesses identify these patterns and they can tailor deals or promotions to take advantage of their consumers' buying habits

Sequential Patterns

- Sequential patterns uses past data to form a predictive model
- Produces projected trends of what the data shows a consumer will buy
  - Example: Target could predict a consumer will buy diapers if they are/have purchased baby clothes and pacifiers in the past

Data Mining’s Major Elements

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Analyze the data by application software.
4. Present the data in a useful format, such as a graph or table, for display and visualization

Methods of Analysis

- Artificial neural networks: Non-linear predictive models that resemble biological neural networks in structure

- Genetic algorithms: Optimization techniques that use genetic combination, mutation, and natural selection based on the concepts of natural evolution

- Rule induction: The extraction of useful if-then rules from data based on statistical significance

Methods of Analysis (Cont.)

- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
  - Classification and Regression Trees (CART), Chi-Square Automatic Interaction Detection (CHAID)
  - CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits.

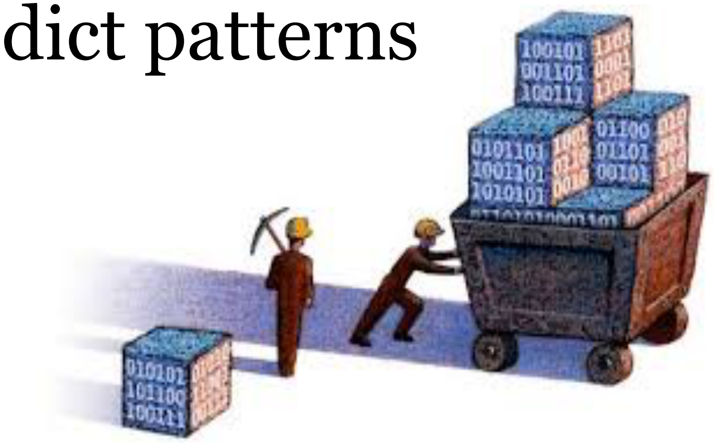
Methods of Analysis (Cont.)

- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k-record(s) most similar to it in a historical dataset (where k=1)

- Data visualization: The visual interpretation of complex relationships in multidimensional data.
  - Use graphics tools to illustrate data relationships

How Data Mining Applied

- Data mining is accomplished through modeling
- Modeling is the act of building a model that applies to one situation and then applying it to another situation where you don’t have a model
- Use these new models to predict patterns



Data Mining and Marketing

- Advances in the data mining field have had profound effects on the marketing of companies
- Companies use this data to tailor their coupons, advertisements and sales to consumers
- This marketing tactic is more effective, efficient and can save the company money

Target Case Study

- Target uses data mining to tailor the coupons they send in hopes to attract consumers at times in their lives where they are vulnerable to changing their store loyalties
  - The period where consumers are most vulnerable is when parents are expecting a child
  - Research has found that when a couple is expecting, they often break their habits and form new ones
  - This gives stores like Target the opportunity to lure consumers into their stores and get them hooked for life

Target (Cont.)

- Target uses data that it collects while you are in the store/on their website along with personal information that they buy from other companies

"For decades, Target has collected vast amounts of data on every person who regularly walks into one of its stores. Whenever possible, Target assigns each shopper a unique code — known internally as the Guest ID number — that keeps tabs on everything they buy" (Duhigg)

- This data is then analyzed to better understand consumers' shopping and personal habits

Target (Cont.)

- Analyst, Andrew Pole, started work on a "pregnancy prediction model" by combing through Target's baby shower registry and taking note of how shopping habits of pregnant women changed throughout their pregnancy
- Using this info, he created a list of about 25 items that signal that a woman is pregnant
- This model was able to predict not only if someone is pregnant but also estimate due date

Amazon Case Study

- Amazon is using the data they have collected to improve the customer-service
  - This includes, name, address & basic personal info as well as consumer preferences and the specific issue the consumer is trying to fix
- Use synchronized data to transfer all the data about an individual collected from various departments to provide the customer service representative with the information they need to have an effective human conversation

Amazon (Cont.)

- It makes interactions with consumers more efficient

- Customer service employees have access to the info needed when interacting with customers

- The employees know enough about you to make your interaction seem personal but not too much that it seems creepy
    - Good to know name, address and the topic of the call but don’t need to suggest an item that your data has shown they may like

Starbucks Case Study

- Starbucks uses data to determine the best locations for their stores

- Multiple Starbucks locations are able to do so well in such close proximity due to data mining and modeling

- Use location-based data, street traffic analysis and demographic information to determine where their locations will have the most success

Starbucks (Cont.)

- Starbucks uses a company called ***Esri*** and their data platform, ***ArcGIS*** online, to monitor sales, demographics and proximity to potential consumers' homes, work and other excursions

- This company takes Starbucks' massive amount of data, analyzes it and places it in easy-to understand platform for Starbucks employees

The Future of Data Mining

- Predictive analytics: "one-click data mining", achieved by a easier and more efficient data-mining process
    - Allow advanced analytics to be applied across subjects
    - The most revolutionary will be in medicine

      Researchers can use predictive analytics to find factors associated with a disease or predict what patient might respond best to an experimental treatment.

Future Trends

- Distributed Data Mining: mining data that is located in various different locations
  - Uses a combination of localized data analysis with a global data model
- Hypertext/Hypermedia Data Mining: mining data which includes text, hyperlinks, text mark-ups, and other forms of hypermedia info
  - Techniques: classification, clustering, semi-structured learning & social network analysis

Future Trends

- Multimedia Data Mining: multimedia data (including images, video, audio and animation) need to be represented differently than traditional data
    - Audio data mining (mining music)

- Spatial/Geographical Data Mining: analyzing info about natural resources, images from satellittes, or topographical data
    - Most of data is image oriented, a lot of it is from different locations

Concerns about Data Mining

Privacy:

- As data mining becomes more widely used, more info is collected about every individual

- Useful applications of this knowledge vs. potentially dangerous misuse

- Possible easy access of data and ill intentions
    - Potential for identify theft and more!

Other Concerns

- User Interface Issues: do visualization tools make the uncovered knowledge interesting & understandable?
- Performance Issues: many analysis tools and statistical methods were designed for smaller sets of data. As the data size increases, how do they scale?
- Trade-off: Do benefits of data collection and data mining outweigh the potential risk?