

Data Mining and Business Intelligence

Amit Basu, PhD

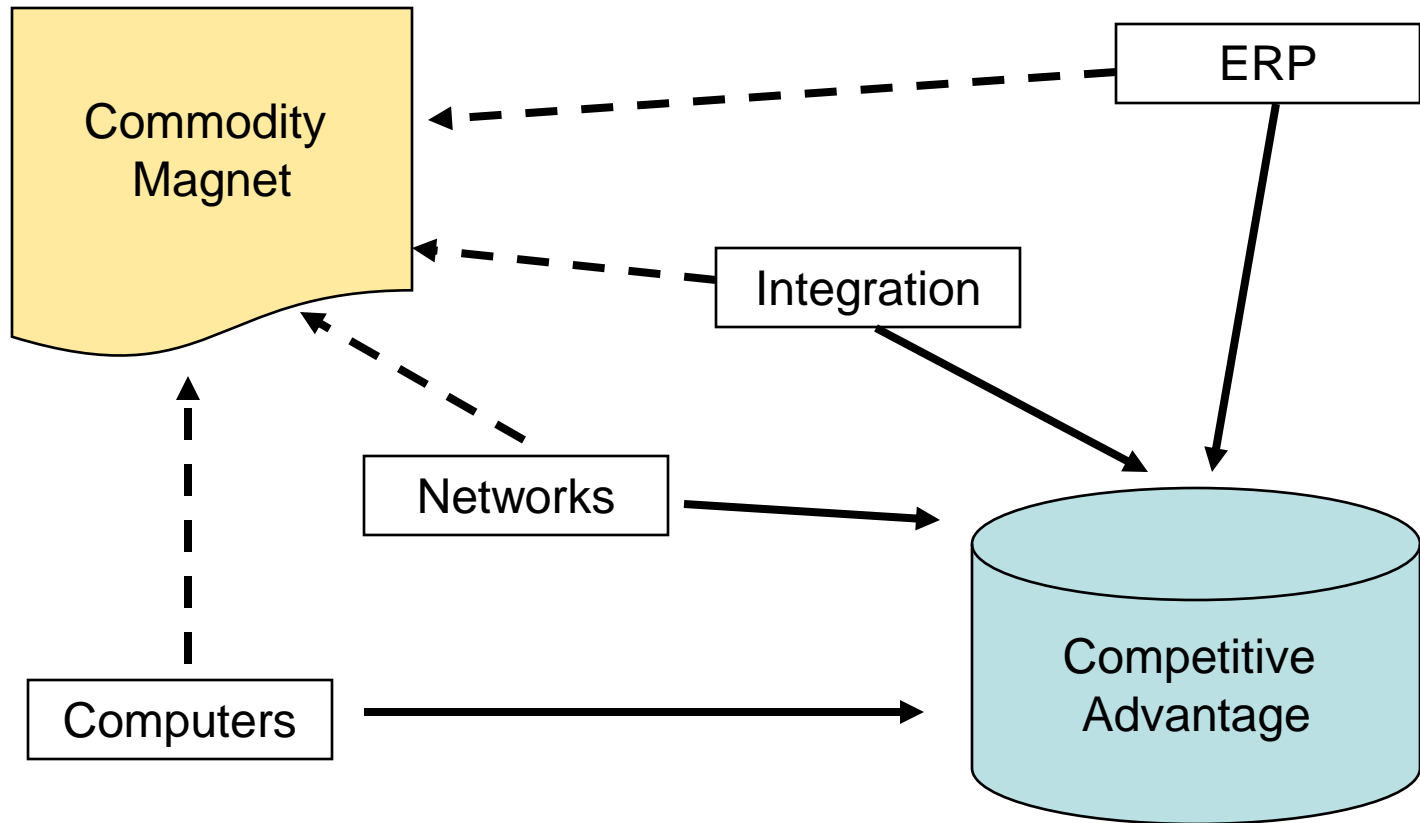
Cox School of Business

SMU

Business Intelligence

- Businesses have invested millions of dollars in IT infrastructure, but the returns have not been clear
- BI is about leveraging this investment effectively
- “Intelligence” as in “Military Intelligence”

The value of IT



Analyzing Data

- Traditional top-down analysis
 - Profitability
 - Efficiency
 - Cost-effectiveness
- Exploratory Analysis
 - What can we discover from our experience?
 - As reflected in our databases
 - Data Mining or “Knowledge Discovery”



Data-driven questions: Customer Profiling

- Who are our biggest customers and what distinguishes them?
- Who are our best customers and how are they different?
- Which customers are changing categories and what distinguishes them?
- How do our prospects differ from our customers?
- How do our customers differ from our competitors' customers?

Contexts and concerns

Example: Banking Applications

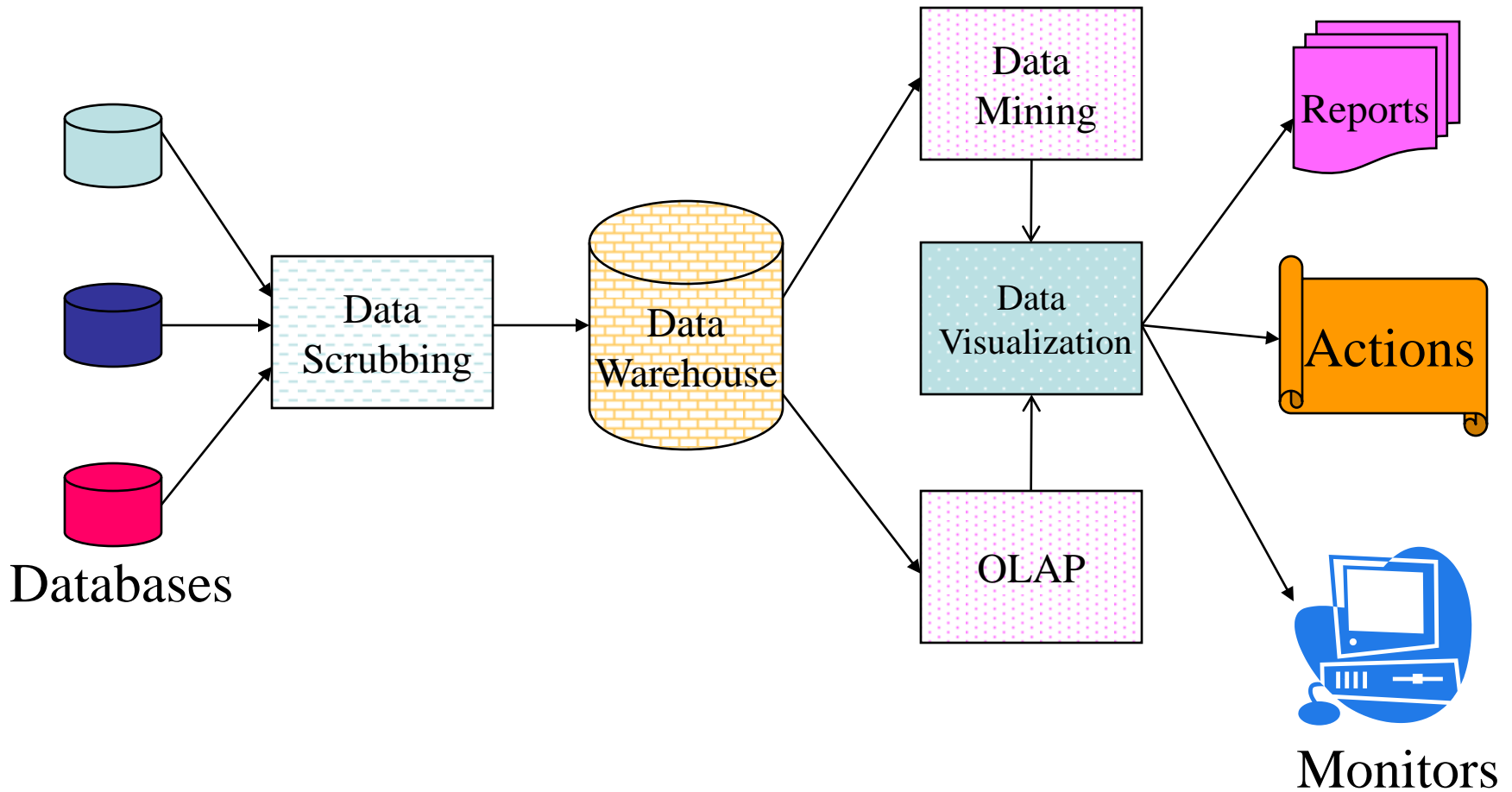
- Corporate Credit Risk Assessment
 - Predicting default – an outlier
 - Data is infrequent and sources are unreliable
 - Rating agencies tend to postpone downgrades
- Consumer Lending
 - Traditionally very crude
 - Based on higher frequency data, which is often incomplete
- Credit Card services
 - Very high frequency data
 - Acquisition, retention, cross-selling, fraud detection, customer service
 - Need for dynamic models due to shifting landscape

Some Basic Questions

- What is Data Mining (DM)?
- Why DM?
- When DM?
- How DM?



Data Mining is a subset of Business Intelligence (BI)



Data Mining – Why now?

1. Data are being produced in huge quantities
 - 11% of all data generated by humanity produced in 1999
2. Data are being stored in accessible formats
3. Computing power is more affordable
4. Competitive pressures are enormous
5. Data Mining software is available

Customer Relationship Management (CRM)



WELCOME TO Your Recommendations

Hello, Ronald Norman. Explore today's featured recommendations. (If you're not Ronald Norman, [click here.](#))

Book Recommendations

Agile and Iterative Development

LOOK INSIDE!



From Book News, Inc.

Larman outlines the principles and best practices of iterative, evolutionary, and agile approaches to software development that emphasize collaboration and flexibility, illustrates those practices in an example system for tracking immigrants, and overviews the work products and core practices of... [Read more](#)

([Why was I recommended this?](#))



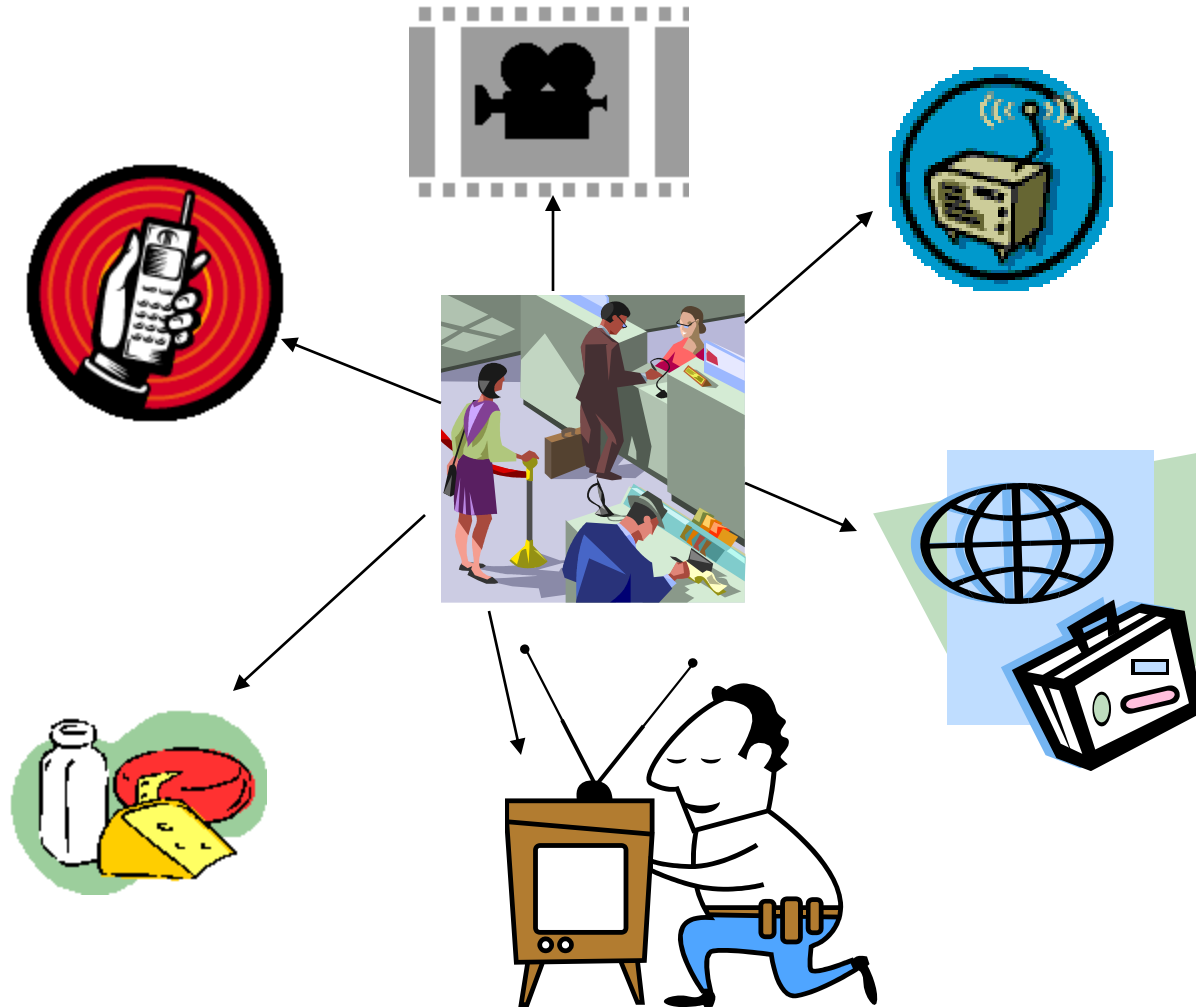
Customer Relationship Management (CRM)

In order to form a learning relationship with its customers, an enterprise (firm) must be able to:

1. **Notice** – what its customers are doing
2. **Remember** – what it and its customers have done over time
3. **Learn** – from what it has remembered
4. **Act On** – what it has learned to make customers more profitable



Identifying and Remembering Relationships is the Key!



Data Mining's Biggest Challenge

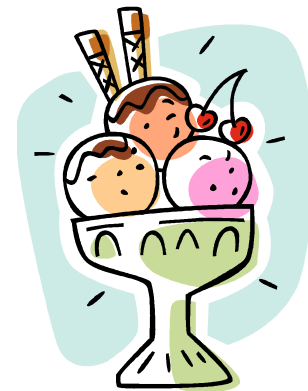


- The largest challenge a data miner may face is the sheer volume of data in the data warehouse.
- A major problem is that this sheer volume may mask the important relationships the data miner is interested in.
 - Granularity of data and model
 - Scope of model – unit of analysis
 - Temporal Relationships

How is Data Mining Different?

- Starts with Data – bottom-up
 - Traditional analysis is model-driven
- **SECONDARY** use of data
 - Relies on data collected for other purposes
 - Data preparation requires substantial effort
- Typically used in dynamic environments
 - Models have to adapt/change over time
- Measures of quality and success are different from traditional methods
 - Confusion matrices and lift charts
- Justification
 - Potential value, not efficiency enhancement

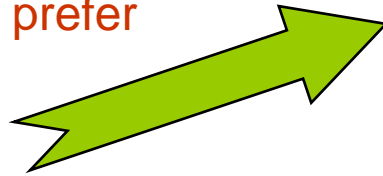
Data Mining Flavors



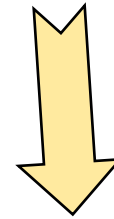
- **Directed** – Attempts to explain or categorize some particular target field such as income or response.
- **Undirected** – Attempts to find patterns or similarities among groups of records without the use of a particular target field or collection of predefined classes.

Data Mining Tasks

- **Description and Profiling**
More guys than gals prefer
BRUT
- **Classification**
Fr, So, Jr, Sr
- **Estimation**
household income
- **Prediction**
predict credit card balance
- **Affinity Grouping**
people who buy X, often buy Y
- **Clustering**



- DW/ OLAP
- Statistics
- Neural Networks
- Genetic Algorithms
- Machine Learning
- Association Rules



- Graphs
- Decision Trees
- Rules
- Relationships
- Templates/Cases

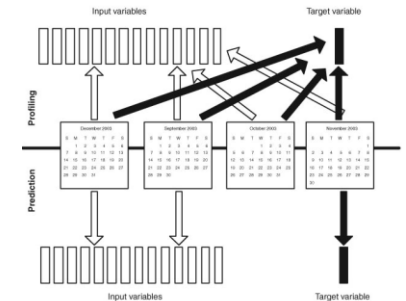
Profiling vs. Prediction

- Profiling

- describes what is in the data
- Demographic variables
- Inability to distinguish cause and effect (eg. Beer drinkers and males)
- Focus is on the past to explain it (timing = past)

- Prediction

- Finding patterns in data from prior period(s) that are capable of explaining or anticipating outcomes in a later period (timing = future)
- Predictive models require separation in time between the model inputs and output.



Learning Things that are not True

- Patterns may not represent any underlying rule
- Sample may not reflect its parent population, hence bias
 - Tanks example
- Data may be at the wrong level of detail (granularity; aggregation)
 - *Who buys the most per month*
 - *Not the most profitable*

Learning Things that are True, but not Useful

- Learning things that are already known
 - *Umbrellas in London*
 - *Married people buy baby-food*
- Learning things that cannot be used
 - *More medicine sales after earthquakes*

The “Dangers” of Data Mining

- Exploitation of the customer
 - How is data about me being used?
- Is it really “knowledge discovery”?
 - *Data dredging*
- He who owns the data owns the knowledge
 - From individual to corporate resource
 - Impact of outsourcing
- Riding on *déjà vu*
 - Volatility and change

DM Examples in Enterprises

- US Government
 - FBI – track down criminals
 - Treasury Dept – suspicious int'l funds transfer
- Phone companies
 - Segmentation, churn analysis
- Supermarkets & Superstores (Vons, Albertsons, Wal-Mart, Costco)
- Mail-Order, On-Line Order (L.L. Bean, Victoria's Secret, Lands End)
- Financial Institutions (B of A, Wells Fargo, Charles Schwab)
- Insurance Companies (USAA, Allstate, State Farm)
- ...

Some Takeaways

- Data Mining should be an integral element of your BI portfolio
 - Can be a means to be creative and bold
 - Fighting the commoditization magnet
- Data Mining is a high-risk, high reward approach
 - Don't bet your job on it
- Effective data mining requires careful planning
 - At the same time, exploratory, iterative process
- Web and mobile data opens up new vistas for data mining