# Data and Text Mining

## Data representation and manipulation II

**prof. dr. Bojan Cestnik**
Temida d.o.o. & Jozef Stefan Institute
Ljubljana
bojan.cestnik@temida.si

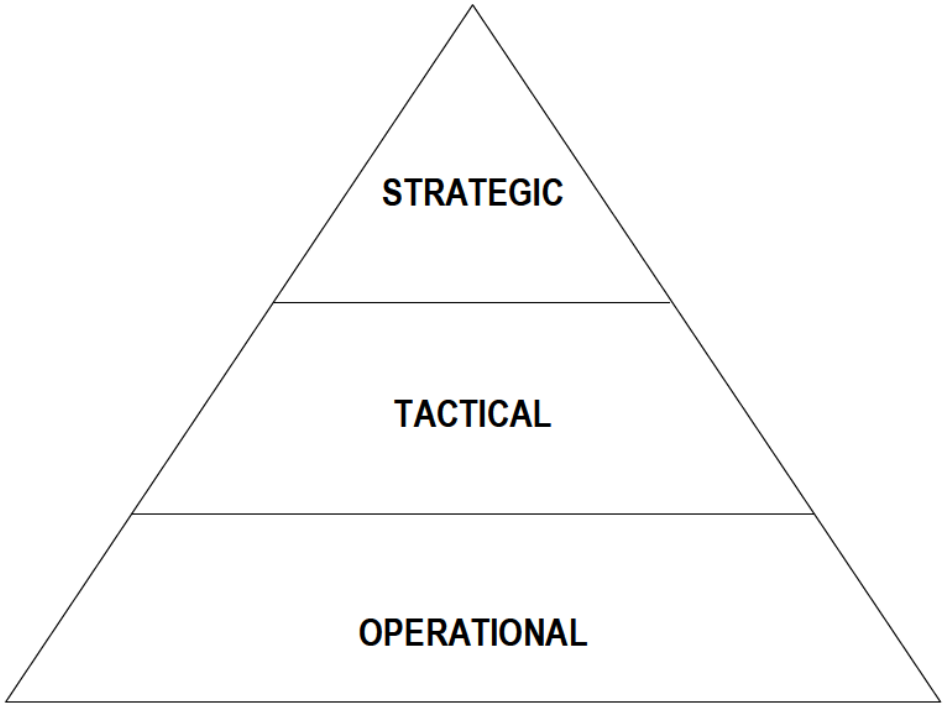
# Contents II

- Supporting decisions with DM
- Data Mining in Marketing
  - CRM – Customer Relationship Management
  - Application areas
  - Examples of practical applications

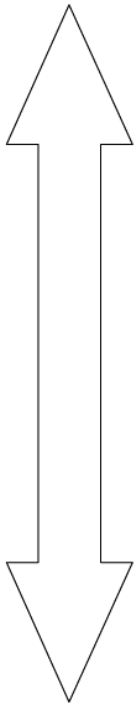
# Good decision-making skills

- One of the skills that sets apart successful business professionals
- Ability to solve complex problems are among the most highly sought after for potential employees
- These skills are directly related to being able to make good decisions
- There is no escaping decision-making

# Types of Decisions




ORGANISATIONAL LEVELS OF MANAGEMENT




STRUCTURE
OF DECISION PROBLEMS

Marko Bohanec

# Why do we make bad decisions?

- Reasons
  - Poor decision making skills
  - Time pressures
  - Relying too much on intuition
  - Overconfident in decision-making skills, intelligence, or knowledge
  - Going with the group
  - Addressing the wrong objective

# Why do we make bad decisions?

- Decision-making biases
  - Negativity bias: giving more weight to negative than to positive experiences
  - Confirmation: searching for info that supports preconceptions
  - Loss aversion: making losses more important than equal gains
  - Bandwagon: tendency to do or believe sth because many others do or believe the same
  - Gambler's fallacy: believing that random events are influenced by previous random events
- Being aware of the biases is the first step in overcoming them

# The cost of spreadsheet errors

- Spreadsheets are widely used analytical tools
- PWC and KPMG estimate that over 90% of spreadsheets contain meaningful errors
- These errors could cost $10 billion per year
- Examples (http://www.eusprig.org/stories.htm):
  - Entering numeric data as text caused a $50K budget short-fall for a UK school
  - The University of Toledo over-projected tuition revenue by over $2M due to an error in a spreadsheet that projected enrollments
  - Incorrect saving of a spreadsheet file misstated natural gas storage amounts, which resulted in inflated prices

# Crisis in classical marketing

- Declining mass markets
- Individual and well informed client
- Limited rational client
- Strong competition
- Traditional marketing approaches can not play a winning role anymore

# Factors for growth and development of new marketing approaches

- Extended globalization
- Stronger degree of competition
- Exacting customers
- Continuously crushing of market segments
- Quickly changing of customers' habits
- Increasing of quality standards
- Technology influence on products and services

(Source: Buttle 2006, Customer relationship management: concepts and tools)

# What is Customer Relationship Management (CRM)?

- CRM is used to learn more about your key customers needs in order to develop a stronger relationship with them
- CRM can be defined as "companies activities related to increasing the customer base by acquiring new customers and meeting the needs of the existing customers"

# CRM characteristics I

- CRM uses technology, strategic planning and personal marketing techniques to **build a relationship** that increases profit margins and productivity
- It uses a business strategy that puts the **customer at the core** of a companies processes and practices

# CRM characteristics II

- CRM brings a change of a companies mindset to become more **customer oriented**
- It requires this customer focused business philosophy to support effective sales, marketing, customer service and order fulfillment
- CRM entails understanding who your customer is and what his specific needs are

# The philosophy of CRM

- The philosophy of CRM is the recognition that your long-term relationships with your customers can be one of the most important assets of an organization, providing **competitive advantage** and improved profitability

(Source: www.it-director.com/)

# DM for CRM I

- "Extraction of hidden predictive information from large databases" (Source: www.thearling.com)
- Powerful new technology with great potential to help companies focus on the most important information in their data warehouses
- DM tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions

# DM for CRM II

- The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems
- Data mining tools can answer business questions that traditionally were too time consuming to resolve

# DM for CRM III

- DM tools scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations
- Most companies already collect and refine massive quantities of data
- DM techniques can be implemented rapidly on existing platforms to enhance the value of existing information resources, and integrated with new products and systems

# Thesis I

- DM could be ideal instrument for managing relationships with clients
- DM tools could help marketers:
  - to find out new knowledge
  - to improve and deepen understanding of customers
  - to transform both together in efficient marketing strategies

# Thesis II

- DM is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:
  - Massive data collection
  - Powerful multiprocessor computers
  - Data mining algorithms
- QUALITY DATA MINING GENERATES NEW BUSINESS OPPORTUNITIES!!!

# DM and CRM I

- The way in which companies interact with their customers has changed dramatically over the past few years - a customer's continuing business is no longer guaranteed
- As a result, companies have found that they need to understand their customers better, and to quickly respond to their wants and needs - the time frame in which these responses need to be made has been shrinking

# DM and CRM II

- It is no longer possible to wait until the signs of customer dissatisfaction are obvious before action must be taken
- To succeed, companies must be proactive and anticipate what a customer desires
- More customers, more products, more competitors, and less time to react means that understanding customers is now much harder to do

# DM and CRM III

- A number of forces are working together to increase the complexity of customer relationships:
  - Compressed marketing cycle times
  - Increased marketing costs
  - Streams of new product offerings
  - Niche competitors

# DM and CRM IV

- Successful companies need to react to each and every one of these demands in a timely fashion
- The market will not wait for your response, and customers that you have today could vanish tomorrow
- Interacting with your customers is also not as simple as it has been in the past

# DM and CRM V

- The need to automate:
  - The Right Offer
  - To the Right Person
  - At the Right Time
  - Through the Right Channel

(Source: www.thearling.com)

# DM tasks

- Classification
- Estimation
- Prediction
- Affinity grouping or association rules
- Clustering
- Description and visualization

# DM applications in marketing

- Improved prospecting
- Better market segmentation
- Increased customer loyalty
- Clearer customer relationship definitions
- More successful cross-selling and up-selling
- Risk management
- More effective and efficient media spending

(Source: www.smartdrill.com)

# Case I – Offering a new product

- Mailing directed at a given customer base
- Typically: 1% of contacted customers are responders who will purchase the offered product
- A mailing of 100,000 will result in about 1,000 sales
- Data mining: identify which customers are most likely to respond to the campaign (based on the past records)
- Response raised from 1% to 1.25%: the sales of 1,000 could be achieved with only 80,000 mailings, reducing the mailing cost by one-fifth

# Case II – Car insurance

- Sports car owners fall into a high-risk category
- By mining driver safety data in data warehouse: if sports car enthusiasts also own a second, conventional car, they may be safe-enough drivers to be attractive policy holders
- As a result of the discovered micro-niche among sports car owners, the company changed how they underwrite and price some sport car policies

# Case III – Customer behavior

- Three types of credit card holders with respect to their profitability:
  - Revolvers: maintain large balance, highly profitable because they pay interest on the balance
  - Transactors: high balance, paid off every month; do not pay interest, just the processing fee
  - Convenience users: periodically charge large bills (vacations, large purchases, …), pay them off several months
- Data: 18 months of billing
- Segmenting by estimating revenue, by potential, by comparison to ideals

# Application areas

- Banking and finance (investment and client analysis, loan approval, …)
- Insurance (client analysis, …)
- Telecommunications (fraud detection, …)
- Retail sales (client analysis, store location and organization, CRM, database marketing, …)
- Medicine (predicting hospitalization costs, discovering diagnostic rules, …)

# The business context for DM

- Application areas:
  - DM as a research tool
  - DM for process improvement
  - DM for marketing (database marketing)
  - DM for Customer Relationship Management (CRM)

# The technical context for DM

- DM and Machine Learning
- DM and Statistics
- DM and Decision Support
  - Data warehouses
  - OLAP, Data marts, Multidimensional databases
- DM and Computer technology

# The societal context for DM

- Individual predictions?
- Open issues:
  - Data ownership?
  - Privacy: a threat or legal obligation?
  - Ethics?

# Why to use DM in marketing?

- The shift of focus from **general observations** (statistics) to **individual descriptions** (DM)

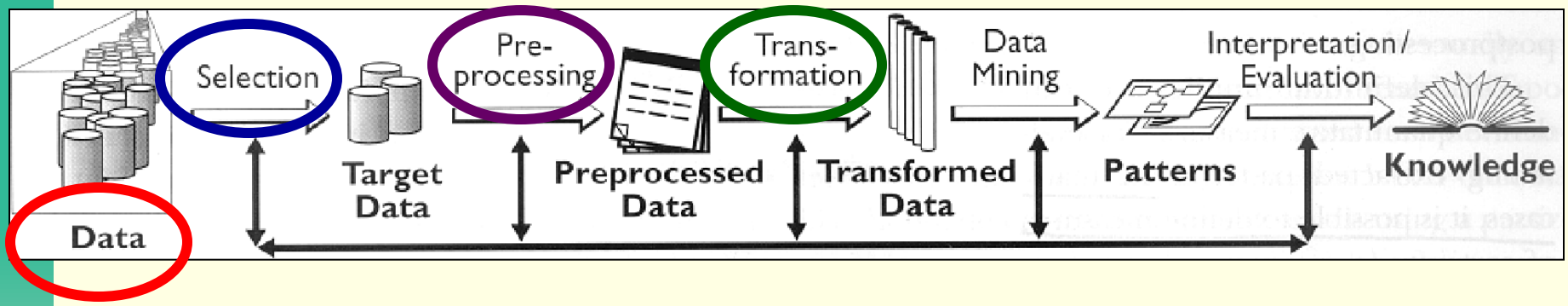
# Four approaches to DM

- Purchasing scores (polaroid camera)
- Purchasing software for a particular application (automated camera)
- Hiring outside experts (wedding photographer)
- Developing in-house expertise (building your own darkroom, becoming a skilled photographer yourself)

# DM methodology

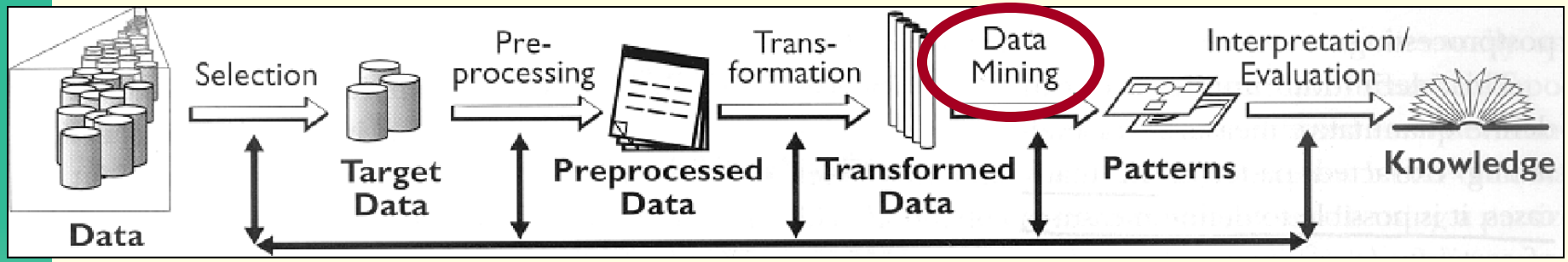
- Two styles:
  - Directed DM – the user knows exactly what s/he wants to predict (model)
  - Undirected DM – the user determines whether the obtained patterns are important

# The process of knowledge discovery from data I




- Tables (n-tuples), relational databases, text, pictures
- Subset selection (data, variables)
- Data cleaning, noise handling, treating missing values
- Transformation in the form required by the algorithms

# The process of knowledge discovery from data II




- Data Mining: the use of the algorithms for data analysis to construct models (rules, decision trees, …) with respect to the task (classification, estimation, prediction, clustering, …)

# Creative cycle of DM

Transform data into actionable information using Data Mining techniques

Act on the information

Measure the results of your efforts to provide insight on how to exploit your data

Identify business problems and areas where analyzing data can provide value

# Identifying business problem

- The trickiest part of successful DM project
- A necessary part of every DM project is talking to the people who understand the business
- Answer questions such as the following:
  - Is the DM effort really necessary?
  - Is there a particular segment that is most interesting?
  - What are the relevant business rules?
  - What do the experts know about the data? Are some data sources known to be invalid? Where should certain data come from?
  - What do expert's intuition and experience say is important?

# Transforming data into results

- Identify and obtain data
- Validate and cleanse the data
- Add derived variables
- Prepare the model set
- Choose the technique and train the model
- Check performance of the models
- Select the most suitable model

# Acting on the results

- Insights
- One-time results
- Remembered results
- Periodic predictions
- Real-time scoring
- Fixing data

# Measuring the effectiveness

- Visualization of the results
- What makes predictive modeling successful?
  - Time frames of predictive modeling
  - Assumptions:
    - The past is a good predictor of the future
    - The data are available
    - The data contain what we want to predict

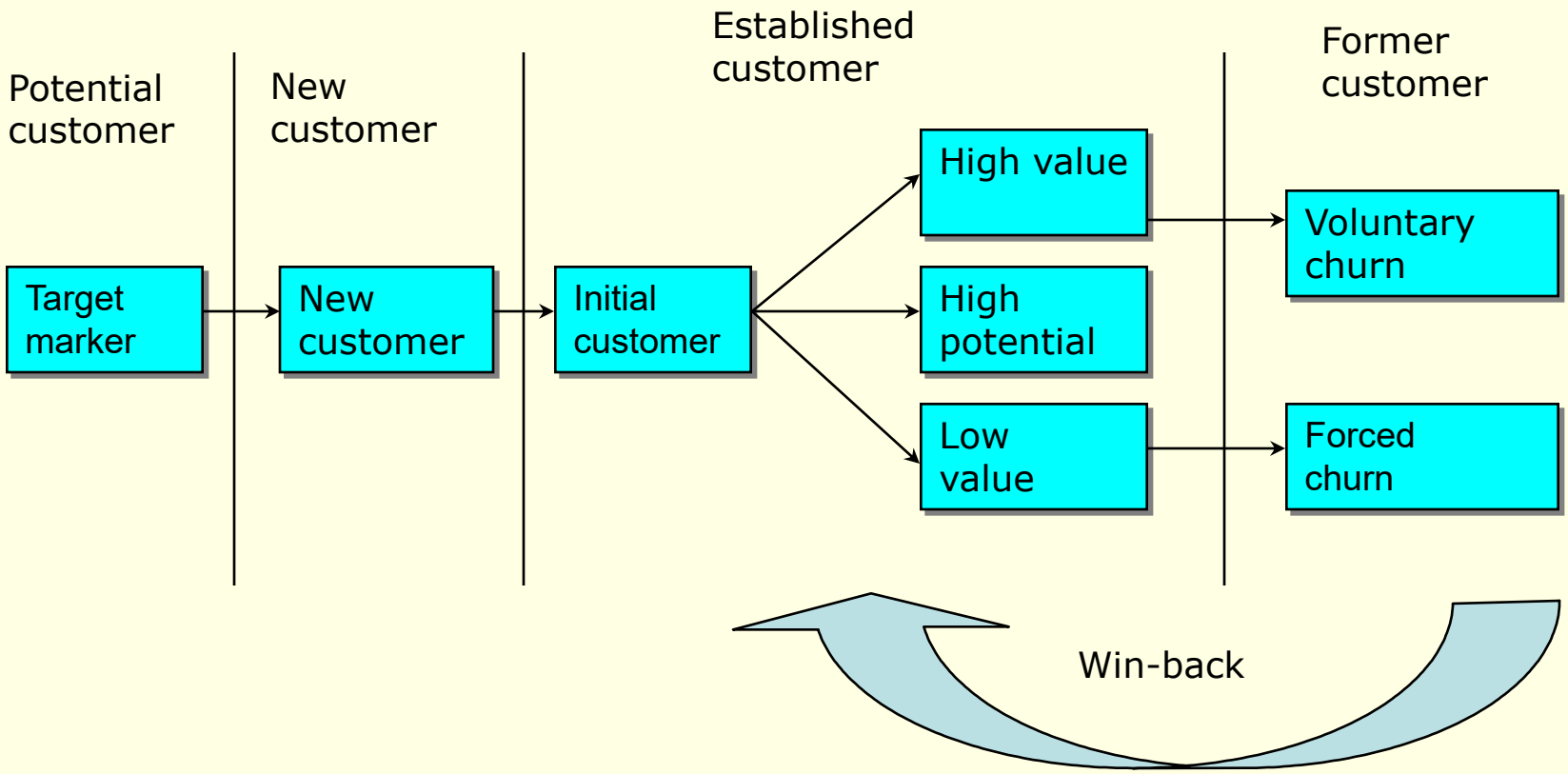
# CRM - Who is the customer?

- Consumer
  - Multiple roles: action role, ownership role, decision-making role
- Business customer
  - Distribution networks
- Customer segments
  - Grouping similar customers (e.g. gold and platinum card holders)

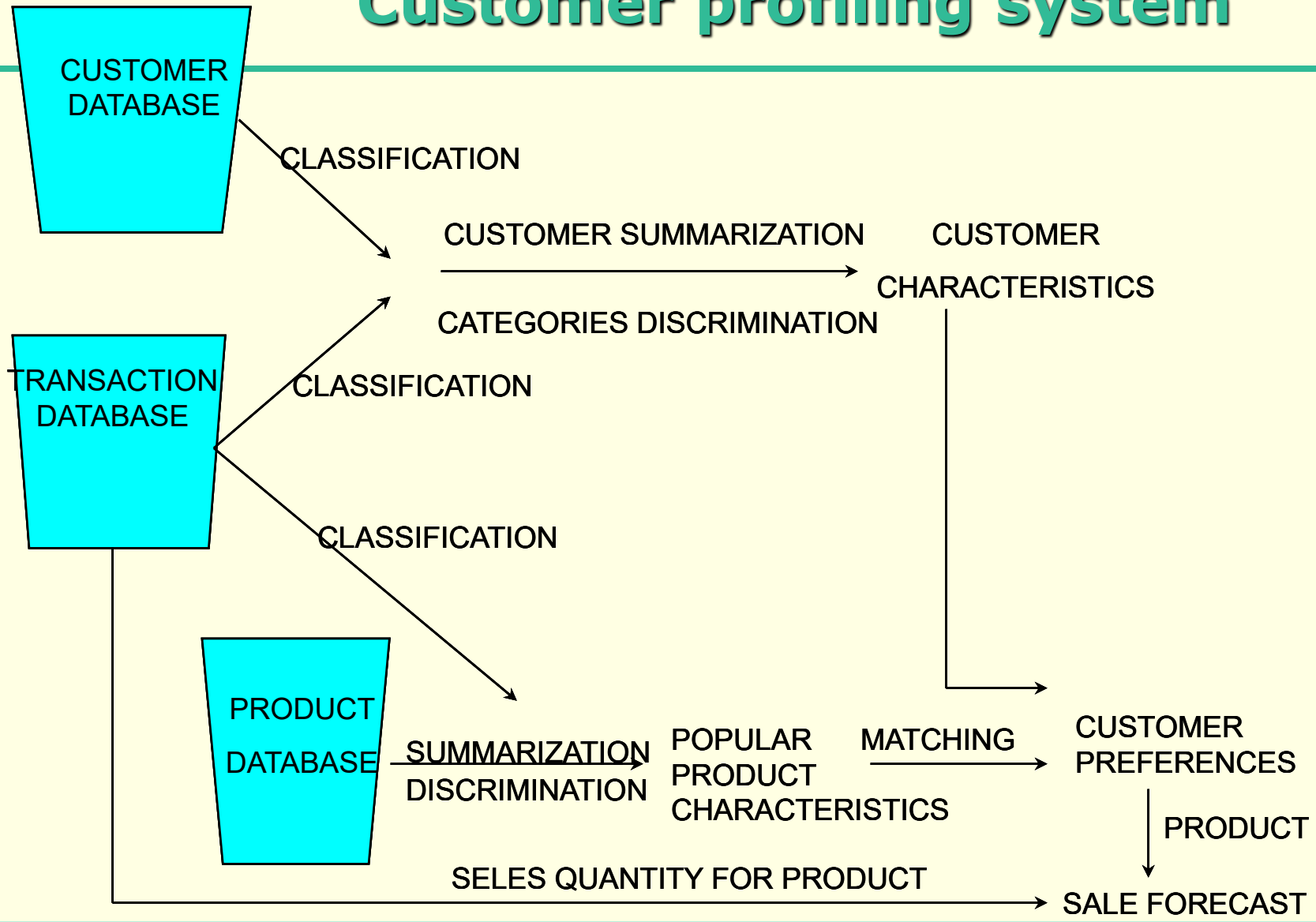
# The customer lifecycle I

- Potential customer
- New customer
- Established customer
  - High value
  - High potential
  - Low value
- Former customer

# The customer lifecycle II



# Customer profiling system



# Profiling clients

- Transaction attitudes that help to get useful client profiles are:
  - Purchasing frequency
  - Purchasing size
  - Last identified purchase
  - Calculating clients value through its life
  - Potential clients
  - (un)success of past marketing campaigns

# Case IV – Churn modeling I

- Churn – a customer of a mobile telephone company that is likely to leave in near future
- The cost of keeping customers around is significantly less than the cost of bringing them back after they leave
- Traditional approach: pick up good customers and persuade them (with a gift) to sign for another year of service
- Data mining: segment the customers, determine what is your value to them, give them what they need (reliability, latest features, better rate for evening calls)
- Consider the timing: finding the optimal point

# Case IV – Churn modeling II

- As a marketing manager for a regional telephone company you are responsible for managing the relationships with the company's cellular telephone customers
- One of your current concerns is customer attention (sometimes known as "churn"), which has been eating severely into your margins
- The cost of keeping customers is less than the cost of bringing them back, so you need to figure out a cost-effective way of doing this

# Case IV – Churn modeling III

- The traditional approach to solving this problem is to pick out your good customers and try to persuade them to sign up for another year of service
- This persuasion might involve some sort of gift (possibly a new phone) or maybe a discount calling plan
- The value of the gift might be based on the amount that a customer spends, with big spenders receiving the best offers

# Case IV – Churn modeling IV

- This solution is probably very wasteful - there are undoubtedly many "good" customers who would be willing to stick around without receiving an expensive gift
- The customers to concentrate on are the ones that will be leaving - don't worry about the ones who will stay

# Case IV – Churn modeling V – value for customer

- This solution to the churn problem has been turned around from the way in which it should be perceived
- Instead of providing the customer with something that is proportional to their value to your company, you should instead be providing the customer with something proportional to your value to them

# Case IV – Churn modeling VI – value for customer

- Give your customers what they need - there are differences between your customers, and you need to understand those differences in order to optimize your relationships
- One big spending customer might value the relationship because of your high reliability, and thus wouldn't need a gift in order to continue with it

# Case IV – Churn modeling VII – value for customer

- A customer who takes advantage of all of the latest features and special services might require a new phone or other gift in order to stick around for another year
- Or they might simply want a better rate for evening calls because their employer provides the phone and they have to pay for calls outside of business hours
- The key is determining which type of customer you're dealing with

# Case IV – Churn modeling VIII – value for customer and timing

- Consider timing in this process - do not wait until a week before a customer's contract and then pitch them an offer in order to prevent them from churning
- By then, they have likely decided what they are going to do and you are unlikely to affect their decision at such a late date
- Don't to start the process immediately - it might be months before they have an understanding of your company's value to them, so any efforts now would also be wasted

# Case IV – Churn modeling IX – value for customer and timing

- The key is finding the correct middle ground, which could very well come from your understanding of your market and the customers in that market
- Viable alternative: *use DATA MINING to automatically find the optimal point*

# Major lifecycle events

- Acquisition campaigns
- Acquisition campaign responses (mail, phone, web page, warranty card, …)
- Initial purchase, following purchases
- Cross-sell campaigns, up-sell campaigns, …
- Forced cancellation, no more usage (silent churn), forced cancellation
- Win-back campaigns or collection campaigns

# Data in the cycle

- Campaign histories, purchased demographics, …
- Product usage, payment history, campaign responses, channel preferences, …
- Termination reasons

# The role of Data Mining

- Wider gap between
  - Ability to collect and store data about customers, products, …
  - Ability to analyze and extract actionable information from the data
- The role of Data mining: bridging the gap