Topics in Business Intelligence

Lecture 1: Introduction to BI & case study

Tommi Tervonen

Econometric Institute, Erasmus School of Economics

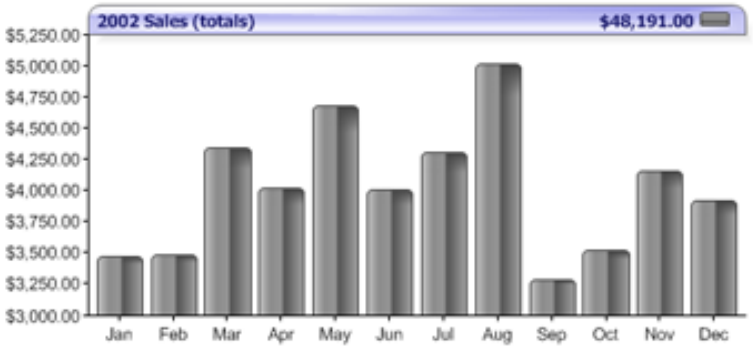
ERASMUS UNIVERSITEIT ROTTERDAM

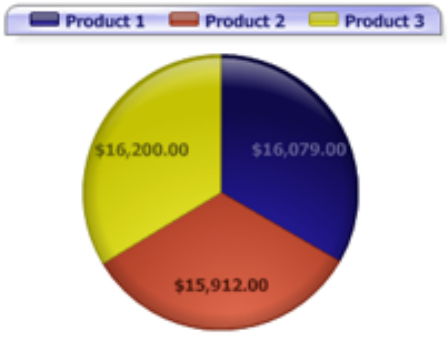
What is Business Intelligence (BI)?

- BI refers to computer-based techniques used in spotting, digging-out, and analyzing business data, such as sales revenue by products and/or departments or associated costs and incomes
- BI technologies provide historical, current, and predictive views of business operations
- Business Intelligence often aims to support better business decision-making

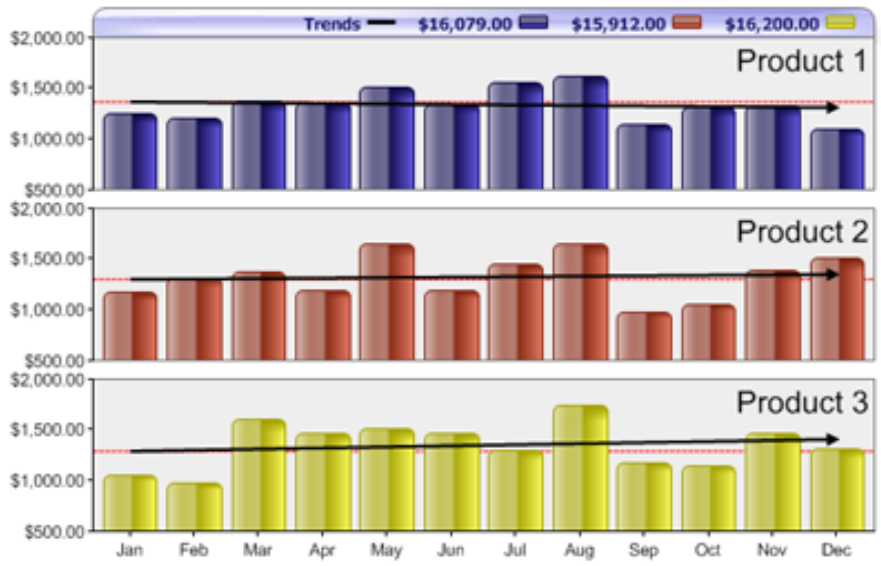
wikipedia.org/wiki/Business_intelligence

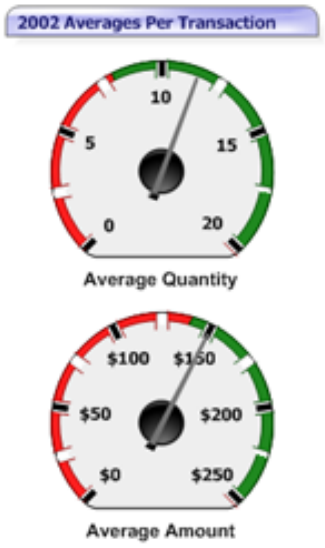
Examples of BI



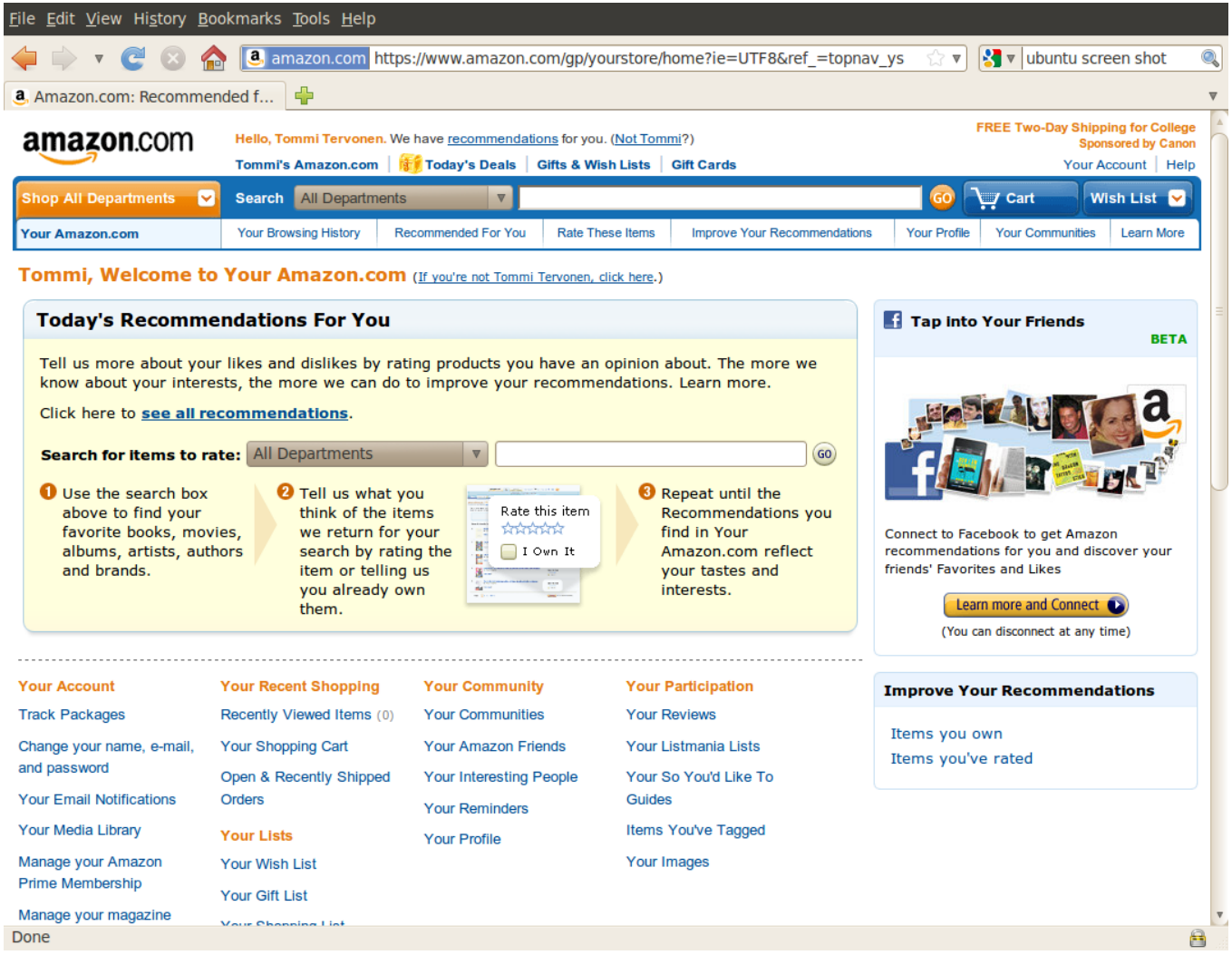


	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	$1,244.00	$1,200.00	$1,374.00	$1,355.00	$1,508.00	$1,348.00	$1,555.00	$1,621.00	$1,137.00	$1,314.00	$1,320.00	$1,103.00
2	$1,167.00	$1,318.00	$1,367.00	$1,191.00	$1,654.00	$1,192.00	$1,451.00	$1,652.00	$976.00	$1,056.00	$1,384.00	$1,504.00
3	$1,058.00	$967.00	$1,600.00	$1,468.00	$1,512.00	$1,470.00	$1,301.00	$1,735.00	$1,172.00	$1,146.00	$1,458.00	$1,313.00
Total	$3,469.00	$3,485.00	$4,341.00	$4,014.00	$4,674.00	$4,010.00	$4,307.00	$5,008.00	$3,285.00	$3,516.00	$4,162.00	$3,920.00



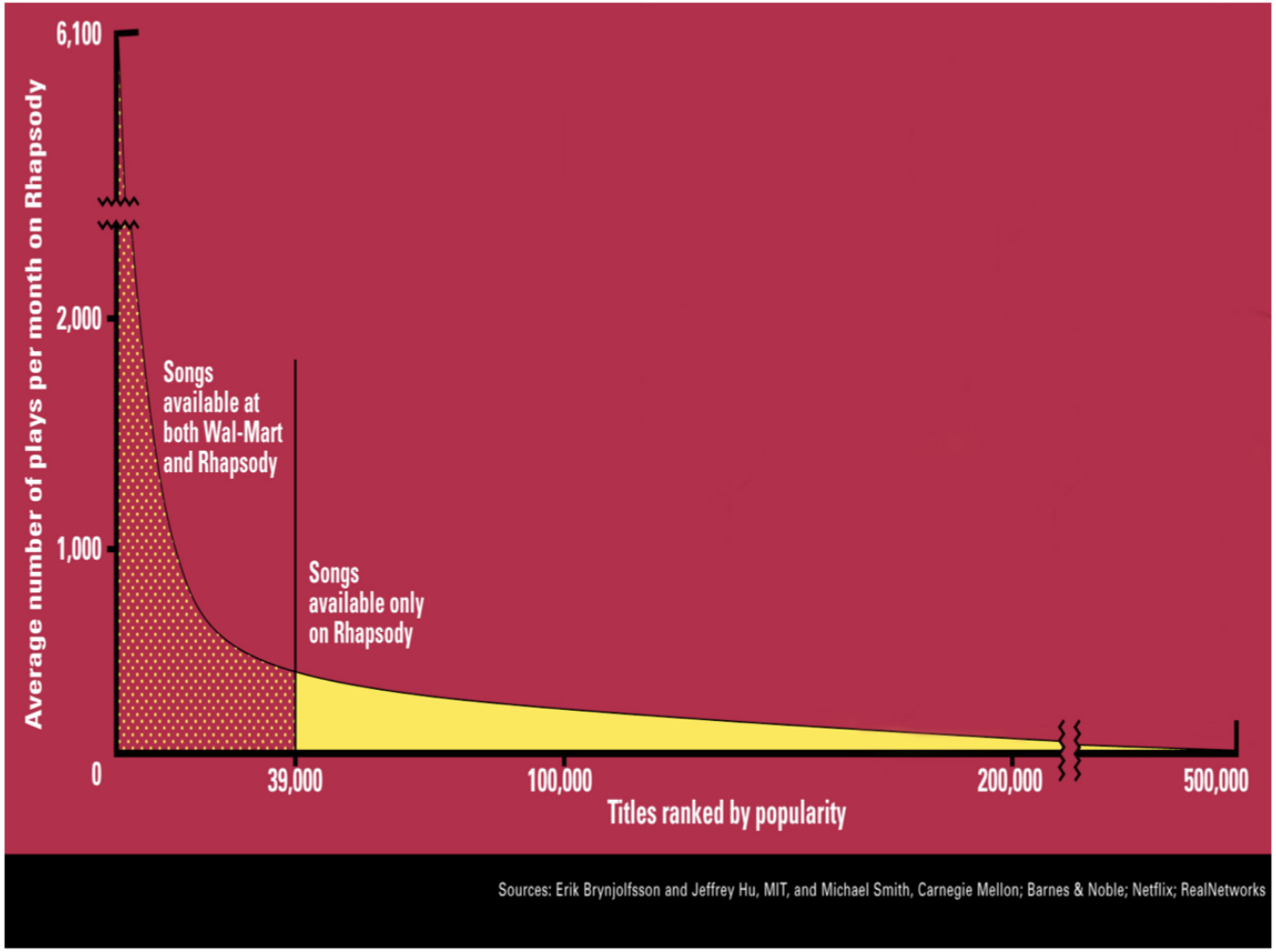


Examples of BI

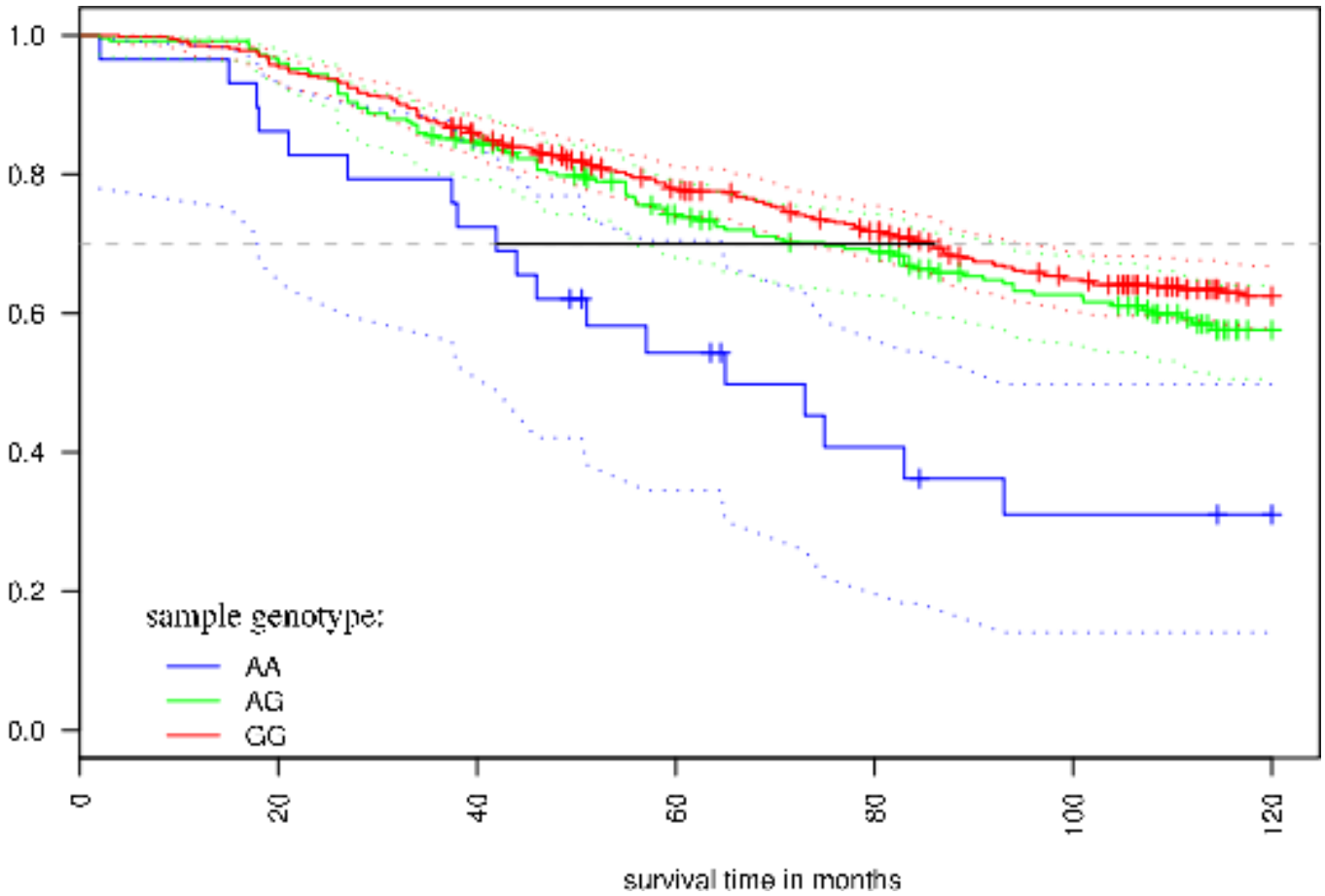


ERASMUS UNIVERSITEIT ROTTERDAM

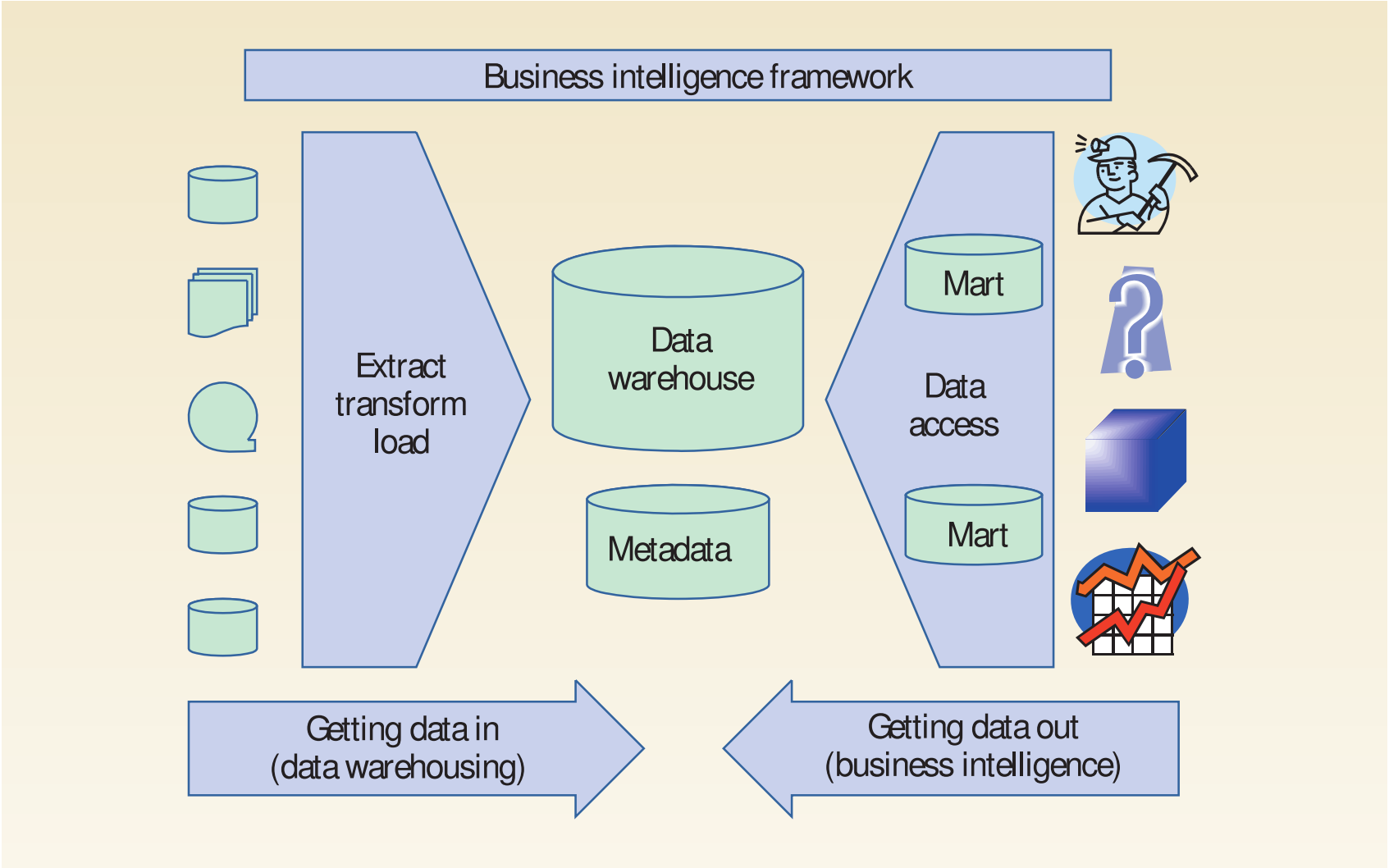
Examples of BI



Examples of BI

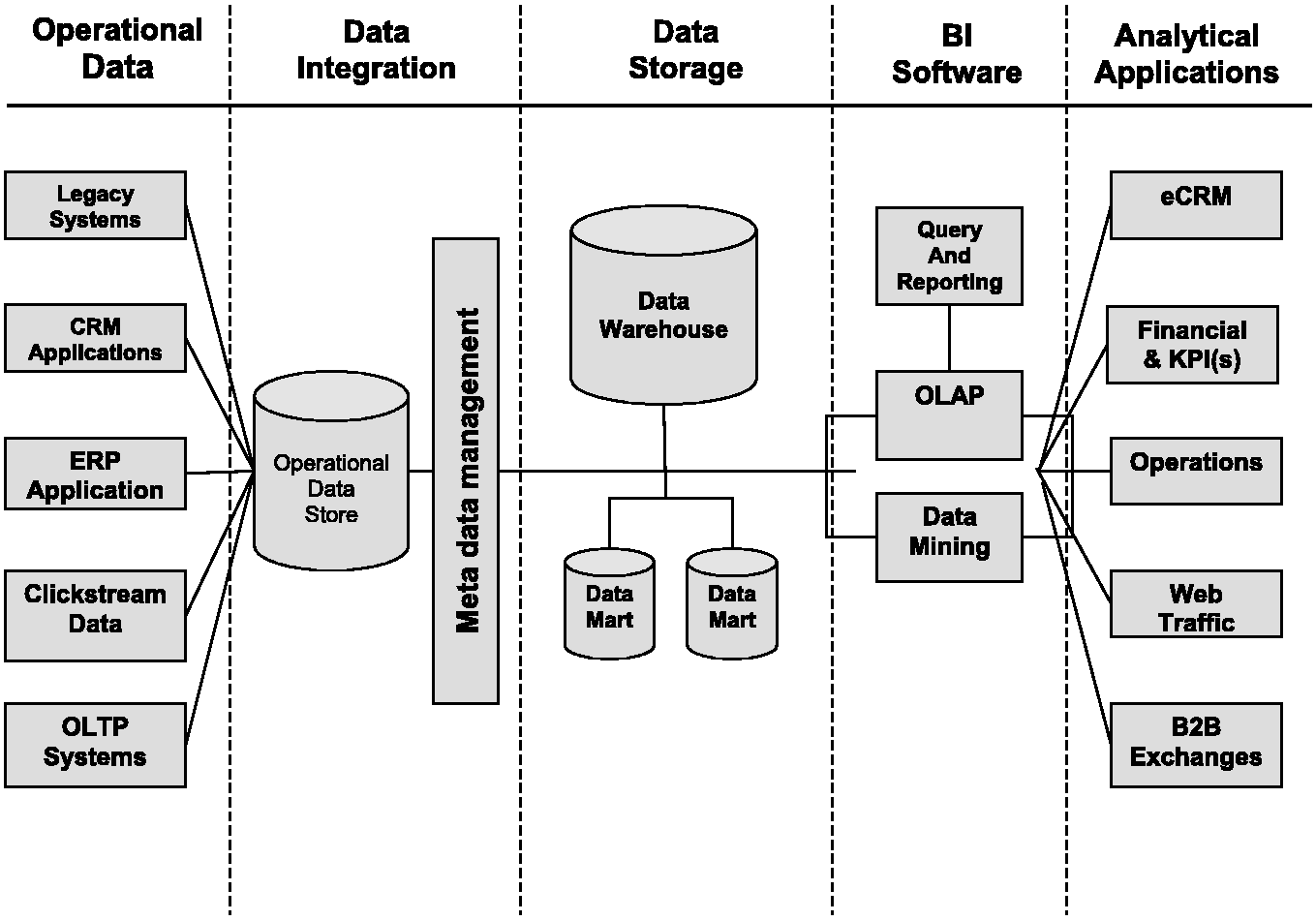


BI framework



Watson & Wixom, 2007

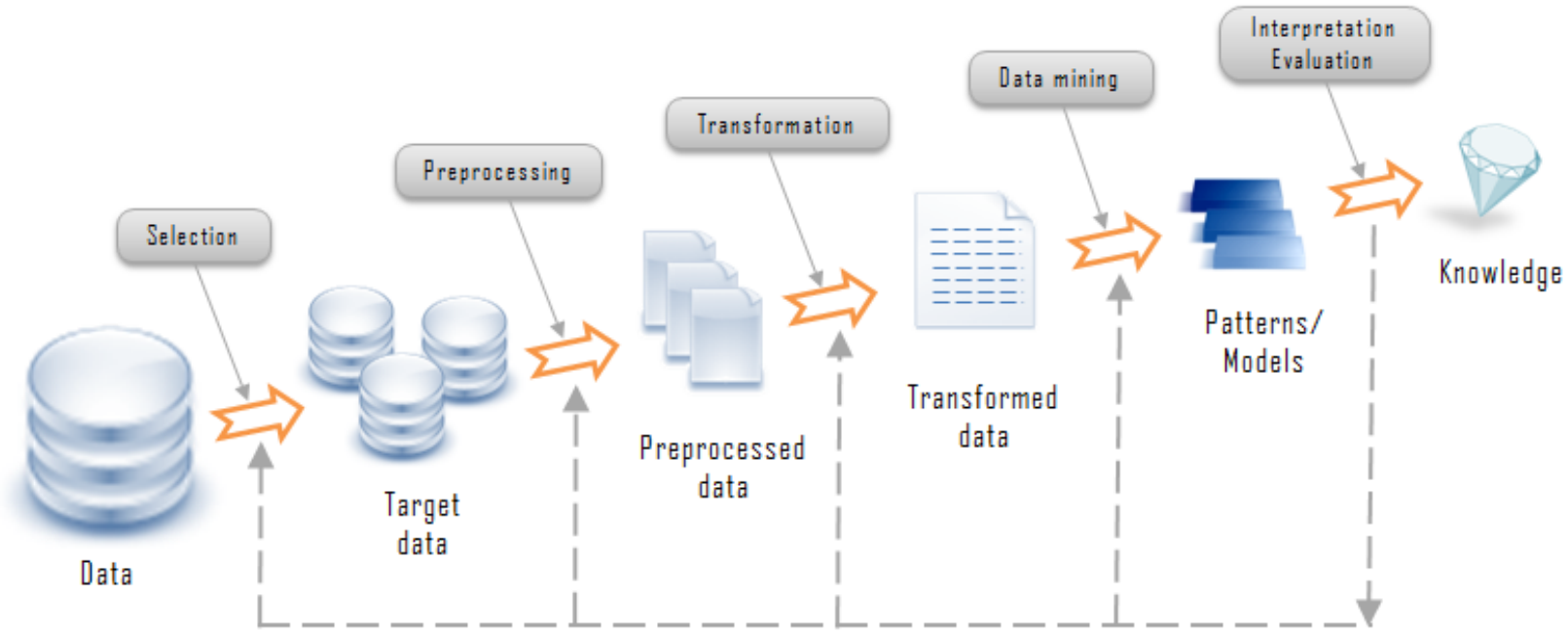
Main components in BI



Source: Datamonitor (2001)

ERASMUS UNIVERSITEIT ROTTERDAM

Knowledge discovery process

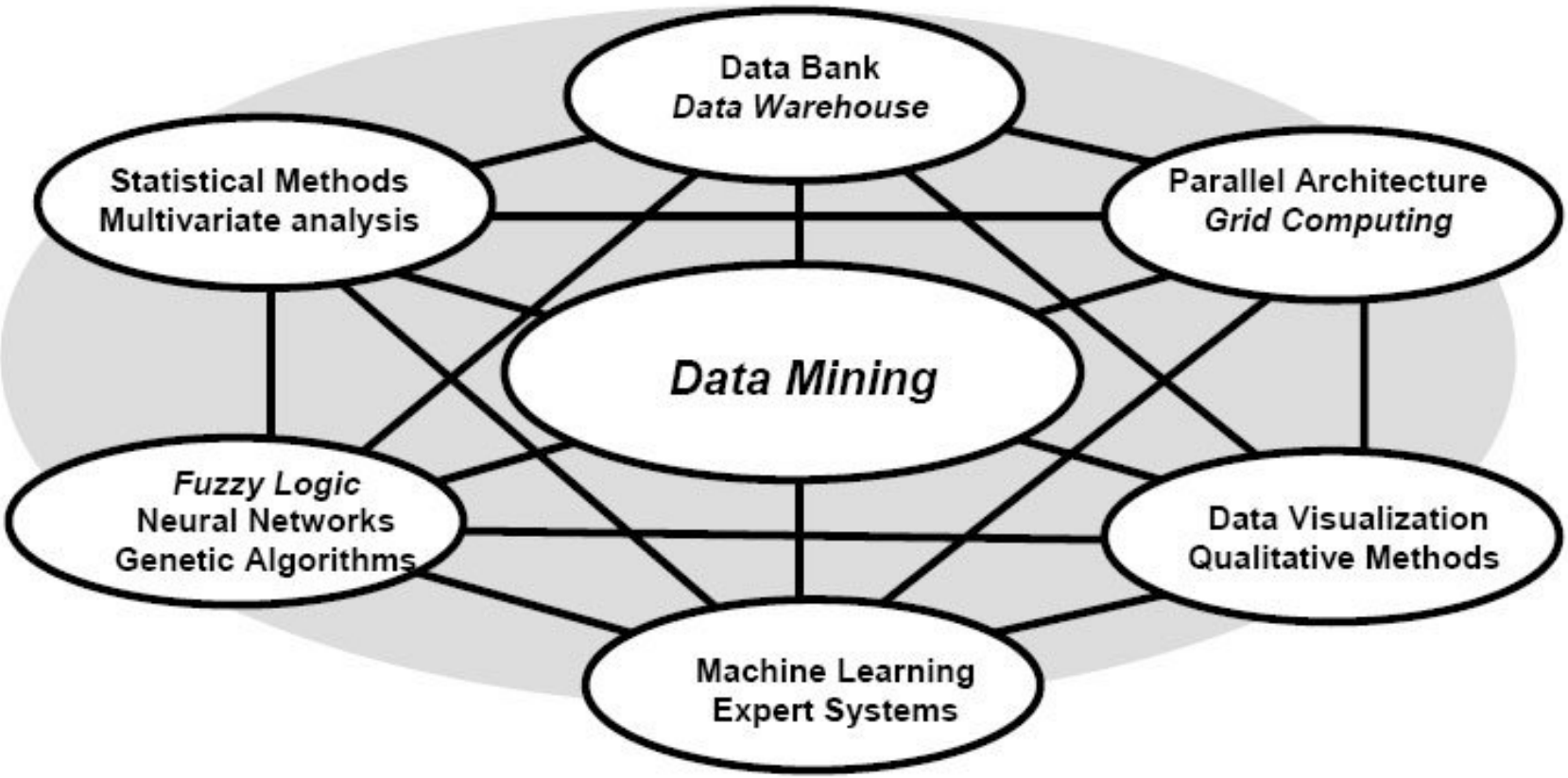


ERASMUS UNIVERSITEIT ROTTERDAM

Why data mining?

- Tremendous amount of data
  - Walmart Customer buying patterns – a data warehouse 7.5 Terabytes large in 1995
  - VISA – Detecting credit card interoperability issues – 6800 payment transactions per second
- High dimensionality of data
  - Many dimensions to be combined together
- High complexity of data
  - Time-series data, temporal data, sequence data
  - Spatial, spatiotemporal, multimedia, text and Web data

Data mining



Subtypes:

- Text mining: mining of patterns from text
- Web mining: discovering patterns from the web

ERASMUS UNIVERSITEIT ROTTERDAM

Data mining: predictive analysis types

- **Classification** of observations to (possibly ordered) classes, e.g. credit card transactions to normal or fraudulent ones
- **Prediction** is similar, but instead of assignment to classes, we try to predict the value of a numerical variable, e.g. amount of credit card purchase
- **Association rules or affinity analysis** tells what is associated with the observations. Recommender systems (e.g. amazon.com) use association rules

Data mining: pre-analyses

- **Data visualization** allows "easy" overview of the data
- **Data exploration** often needs to be done with large data sets to answer more vague questions. Similar variables and observations can be aggregated to get a better picture of the data
- **Data reduction** consolidates a large number of variables or cases into a smaller set. Correlation & principal component analyses

What is 'data'?

- Data can essentially be:

  1. Continuous – ordered values with a scale. E.g. client monthly spending (€), speed of car (km/h)

  2. Categorical – discrete, possibly ordered values. E.g. car class (small family car, large family car, executive, ...), bank customer credit class (A, B, C, D)

Often data is categorical due to form of reporting (e.g. from questionnaires: monthly salary)

Data mining methods for BI

Mostly:

- Statistical methods for analysis of continuous variables
- Machine learning for analysis of categorical variables
- Variables are divided into predictors and responses

Data nature & methods

	Continuous response	Categorical response	No response
Continuous predictors	Linear regression Neural nets *k*-nearest neighbors	Logistic regression Neural nets Discriminant analysis *k*-nearest neighbors	Principal components Cluster analysis
Categorical predictors	Linear regression Neural nets Regression trees	Neural nets Classification trees Logistic regression Naive Bayes	Association rules

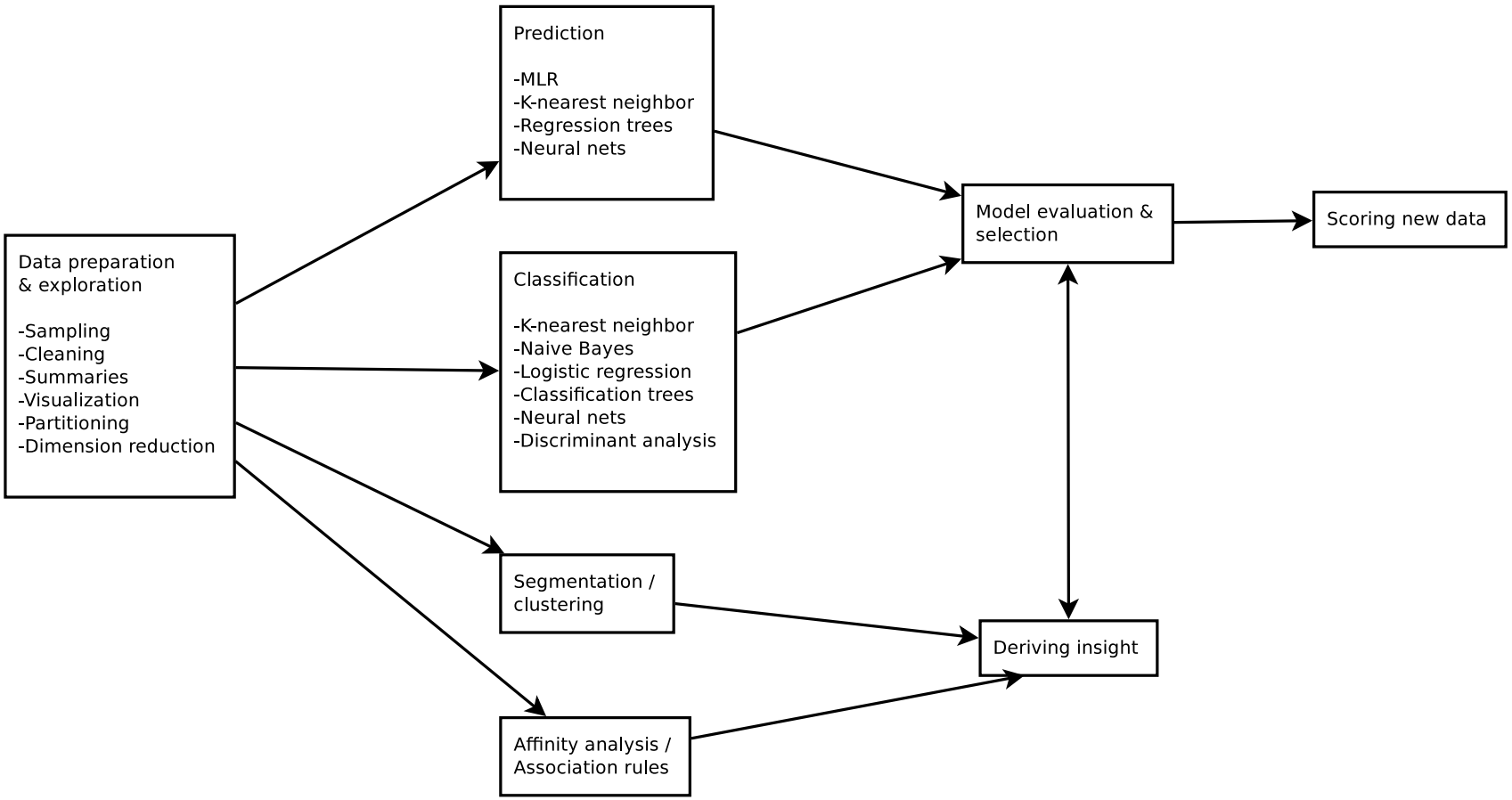
- Ordered categorical variables (e.g. 1, 2, 3) can often be converted to continuous ones
- Continuous variables can always be converted to categorical ones through frequency analysis (binning)

Data nature & methods

	Continuous response	Categorical response	No response
Continuous predictors	Linear regression Neural nets *k*-nearest neighbors	Logistic regression Neural nets Discriminant analysis *k*-nearest neighbors	Principal components Cluster analysis
Categorical predictors	Linear regression Neural nets Regression trees	Neural nets Classification trees Logistic regression Naive Bayes	Association rules

- Ordered categorical variables (e.g. 1, 2, 3) can often be converted to continuous ones
- Continuous variables can always be converted to categorical ones through frequency analysis (binning)

Data mining process



Learning modes

- In **unsupervised learning**, no outcome variable is predicted.

Segmentation / clustering

Affinity analysis / Association rules

Learning modes

- In **supervised learning** the model is trained to predict a known response
- The data needs to be split into **training** and **test** sets

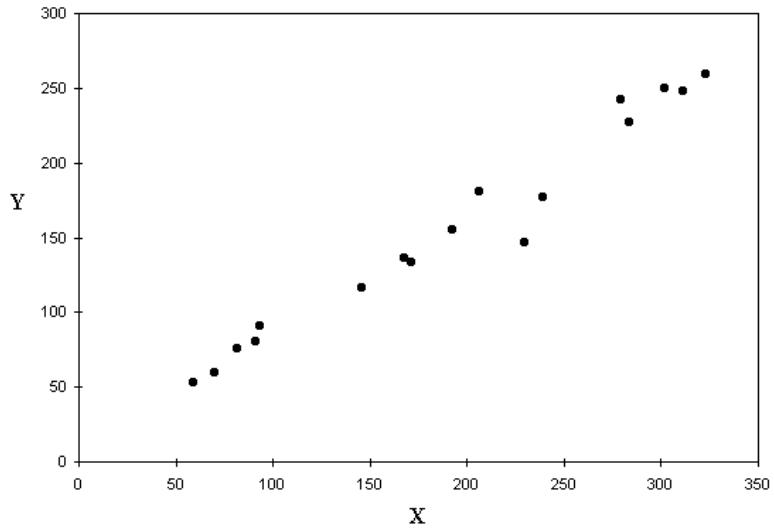
Prediction

-MLR
-K-nearest neighbor
-Regression trees
-Neural nets

Classification

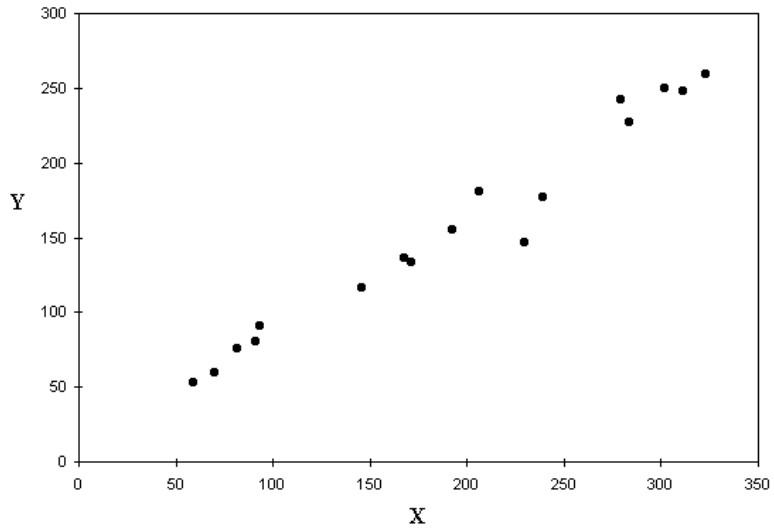
-K-nearest neighbor
-Naive Bayes
-Logistic regression
-Classification trees
-Neural nets
-Discriminant analysis

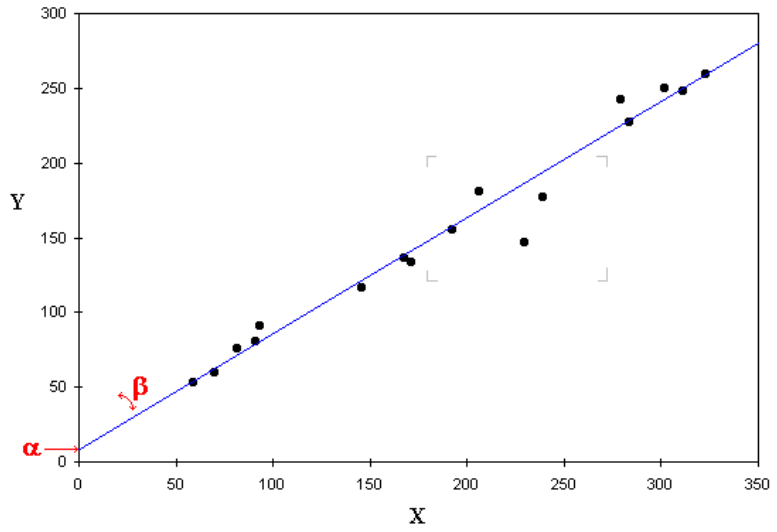
Supervised learning with linear regression



ERASMUS UNIVERSITEIT ROTTERDAM

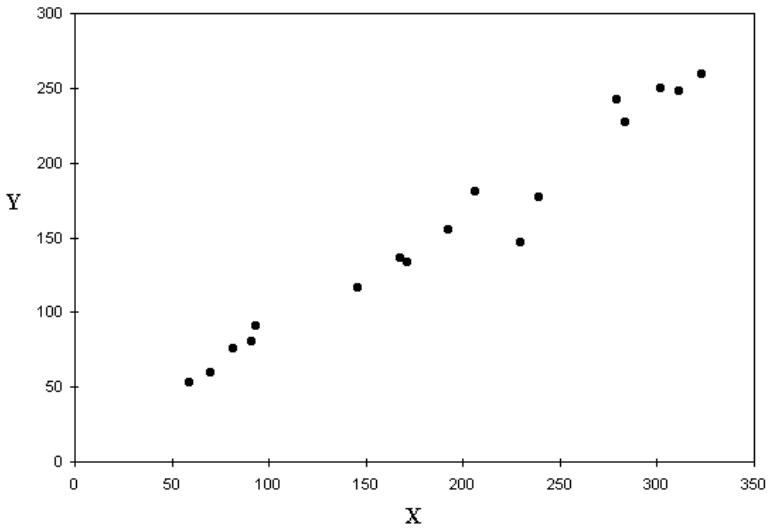
Supervised learning with linear regression

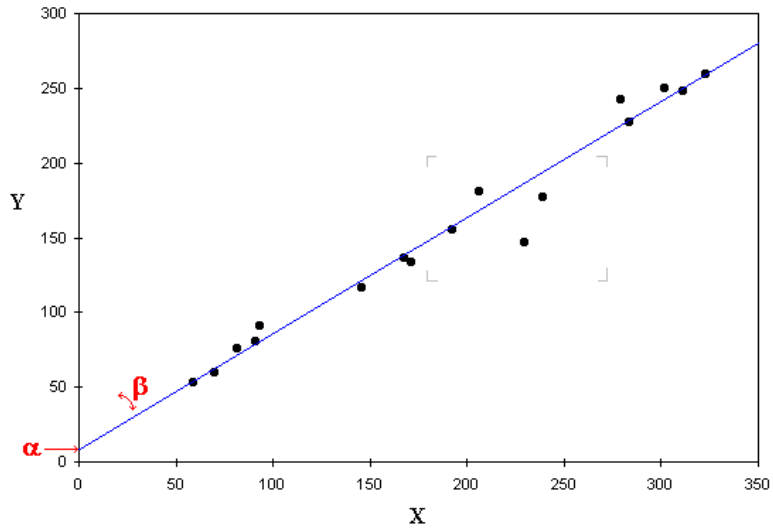




ERASMUS UNIVERSITEIT ROTTERDAM

Supervised learning with linear regression





$x = 200,\ y = ?$

ERASMUS UNIVERSITEIT ROTTERDAM

Data mining process

1. Develop an understanding of the purpose of the data mining project
2. Obtain the dataset to be used in the analysis
3. Explore, clean, and preprocess the data
4. Reduce the data, if necessary, and (in supervised learning) separate into training, test, and validation sets
5. Determine the data mining task (classification, prediction, etc)
6. Choose the technique to be used
7. Apply algorithms
8. Interpret results
9. Deploy model

Q?

Course organization

Lectures:

1st Introduction to BI & case study

2nd Data reduction

3rd Model validation

4th No lecture - use the time for preparing your presentation

5th Student lectures: Naive Bayes and k-NN, Classification trees

6th Student lectures: Logistic regression, Neural nets

7th Overview of results, comparison with (yet another) test set, feedback

Course learning objectives

1. Knowledge of basic principles of data warehouse
2. Comprehension of business implications of BI and data mining
3. Application of a single data mining classification method
4. Evaluation of data mining results

Course evaluation & material

Evaluation:

- Student lecture & case analysis (100%)
- Student lectures have mandatory attendance (1 miss allowed)

Course evaluation & material

Evaluation:

- Student lecture & case analysis (100%)
- Student lectures have mandatory attendance (1 miss allowed)

Online material (all will be available @ `http://smaa.fi/tommi/courses/tbi/`):

- My slides from the first 3 lectures
- Slides of the student lectures
- Scientific papers

Course book: Shmueli, Patel & Bruce, "Data mining for Business Intelligence" - helps in making the student lecture but is not mandatory

Student lectures

- Prepared in pairs or small groups
- Each lecture should consist at least the following:
  1. Theoretical explanation of the method
  2. An application of the method to a simple case
  3. Presentation of real-life BI applications of the method
  4. Analysis of the case study with the method
- Each lecture should be 40mins + 5min discussion: expect to spend 2 weeks in preparation

Case study

- Direct mailings to potential customers ("junk mail") can be an effective way to market a product of service. However, most junk mail is of no interest to majority of people, and ends up being thrown away

Case study

- Direct mailings to potential customers ("junk mail") can be an effective way to market a product of service. However, most junk mail is of no interest to majority of people, and ends up being thrown away

- More directed marketing to highly potential customers saves waste & effort, and consequently lowers costs and increases profits

Case study scope

- Our customer is a Dutch charity organization that wants to be able to **classify** it's supporters to donators and non-donators. The non-donators are sent a single marketing mail a year, whereas the donators receive multiple ones (up to 4).

Case study scope

- Our customer is a Dutch charity organization that wants to be able to **classify** it's supporters to donators and non-donators. The non-donators are sent a single marketing mail a year, whereas the donators receive multiple ones (up to 4).

- Tasks:

  1. Develop a data mining model for classifying the customers to donators and non-donators

  2. Explain through the model which factors are important in deciding who is a donator

Case study data

- Information about donators in 8 variables:

TIMELR	TIME since Last Response (nr weeks)
TIMECL	TIME as CLient (nr years)
FRQRES	FReQuency of RESponse (to mailings)
MEDTOR	MEDian of Time Of Response
AVGDON	AVeraGe DONation (per responded mailing)
LSTDON	LaST DONation
ANNDON	Average ANNual DONation
DONIND	Donation indicator in the considered mailing (response)

- Training and test sets of over 4000 customers

Tools

- Spreadsheet software (e.g. gnumeric, OpenOffice calc, or Excel)

- RapidMiner: an open-source, cross platform tool with available commercial support

Motivation: current directions in BI (debatable)

- **Packaged analytic applications** delivered as both on premises software and software as a service (SaaS) will push control of the information used for decision making toward business units and away from IT organizations
- The economic crisis will reveal which enterprises have a sound information infrastructure and which do not
- The application of social software to the collaborative decision making process will demonstrate the business value of the information coming from BI systems by directly tying it to decisions made

Gartner Inc., 2009

Rhine's paradox

- Joseph Rhine was a parapsychologist in the 1950's who hypothesized that some people had Extra-Sensory Perception
- He devised an experiment where subjects were asked to guess 10 hidden cards red or blue

Rhine's paradox

- Joseph Rhine was a parapsychologist in the 1950's who hypothesized that some people had Extra-Sensory Perception
- He devised an experiment where subjects were asked to guess 10 hidden cards red or blue
- He discovered that almost 1 in 1000 had ESP – they were able to get all 10 right!

Rhine's paradox

- He told these people they had ESP and called them in for another test of the same type
- Alas, he discovered that almost all of them had lost their ESP
- What did he conclude?

You should't tell people they have ESP.

It causes them to lose it.

Bonferroni's principle

- "If you look for interesting patterns in more places than your amount of data will support, you are bound to find crap"

1st week of case study

(Download, install, and explore RapidMiner)

1. Develop an understanding of the purpose of the data mining project
2. Obtain the dataset to be used in the analysis
3. Explore the data

(Import data into RapidMiner)