Topics in Business Intelligence
Lecture 1: Introduction to BI & case study

Tommi Tervonen
Econometric Institute, Erasmus University 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
Examples of BI

Tommi, Welcome to Your Amazon.com

Today's Recommendations For You

Tell us more about your likes and dislikes by rating products you have an opinion about. The more we know about your interests, the more we can do to improve your recommendations. Learn more.

Click here to see all recommendations.

1. Use the search box above to find your favorite books, movies, albums, artists, authors and brands.
2. Tell us what you think of the items we return for your search by rating the item or telling us you already own them.
3. Repeat until the Recommendations you find in Your Amazon.com reflect your tastes and interests.

Your Account
Track Packages
Change your name, e-mail, and password
Your Email Notifications
Your Media Library
Manage your Amazon Prime Membership
Manage your magazine

Your Recent Shopping
Recently Viewed Items (0)
Your Shopping Cart
Open & Recently Shipped Orders
Your Lists

Your Community
Your Communities
Your Amazon Friends
Your Interesting People
Your Reminders
Your Profile

Your Participation
Your Reviews
Your Listmania Lists
Your So You'd Like To Guides
Items You've Tagged
Your Images

Improve Your Recommendations
Items you own
Items you've rated
Examples of BI

Sources: Erik Brynjolfsson and Jeffrey Hu, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks
Examples of BI

sample genotype:
- AA
- AG
- GG

survival time in months
Knowledge discovery process

1. Data
2. Selection
3. Preprocessing
4. Transformation
5. Data mining
6. Patterns/Models
7. Interpretation/Evaluation
8. Knowledge
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
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 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.
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
<table>
<thead>
<tr>
<th>Continuous predictors</th>
<th>Continuous response</th>
<th>Categorical response</th>
<th>No response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>Logistic regression</td>
<td>Neural nets</td>
<td>Principal components</td>
</tr>
<tr>
<td>Neural nets</td>
<td>Discriminant analysis</td>
<td>$k$-nearest neighbors</td>
<td>Cluster analysis</td>
</tr>
<tr>
<td>$k$-nearest neighbors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Categorical predictors</th>
<th>Continuous response</th>
<th>Categorical response</th>
<th>No response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>Neural nets</td>
<td>Association rules</td>
<td></td>
</tr>
<tr>
<td>Neural nets</td>
<td>Classification trees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression trees</td>
<td>Logistic regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive Bayes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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)
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.

<table>
<thead>
<tr>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
</tr>
<tr>
<td>K-nearest neighbor</td>
</tr>
<tr>
<td>Regression trees</td>
</tr>
<tr>
<td>Neural nets</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-nearest neighbor</td>
</tr>
<tr>
<td>Naive Bayes</td>
</tr>
<tr>
<td>Logistic regression</td>
</tr>
<tr>
<td>Classification trees</td>
</tr>
<tr>
<td>Neural nets</td>
</tr>
<tr>
<td>Discriminant analysis</td>
</tr>
</tbody>
</table>
Supervised learning with linear regression

\[ x = 200, \, y = ? \]
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
Lectures:

1st  Introduction to BI & case study
2nd  Data reduction
3rd  Model validation
4th  Student lecture: Naive Bayes and $k$-NN, Classification trees
5th  Student lecture: Logistic regression
6th  Student lecture: Neural nets
7th  Overview of results, comparison with (yet another) test set, feedback
Course learning objectives

1. Knowledge of basic principles of data warehouses
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
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 or service. However, most junk mail is of no interest to the 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.
Our customer is a Dutch charity organization that wants to be able to **classify** its 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
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.
- He discovered that almost 1 in 1000 had ESP – they were able to get all 10 right!
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 shouldn’t tell people they have ESP.

It causes them to lose it.
“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)