

Tomáš Horváth

BUSINESS ANALYTICS

Lecture 0

Overview

Information Systems and Machine Learning Lab

University of Hildesheim

Germany



“Business analytics is the continuous iterative **exploration and investigation of past** business performance **to** gain insight and **drive** business **planning**.” (Michael Benner & Alan Barnett)¹

“I think of analytics as a subset of BI based on **statistics, prediction and optimization**. The great bulk of BI is much more focused on reporting capabilities. ‘Analytics’ has become a sexier term to use – and it certainly is a sexier term than ‘reporting’ – so it’s slowly replacing ‘BI’ in many instances” (Tom Davenport)²

¹ <http://www.docstoc.com/docs/7486045/Next-Generation-Business-Analytics-Technology-Trends>

² <http://www.informationweek.com/news/software/bi/showArticle.jhtml?articleID=222200096>

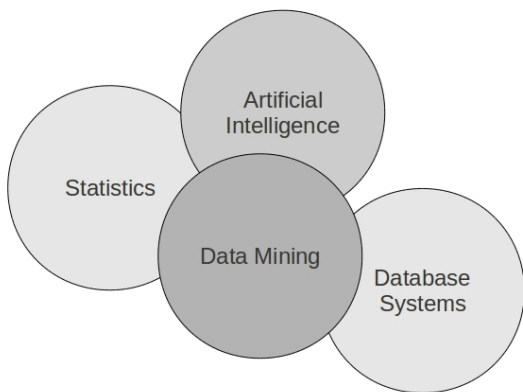
What is the course about?

First of all, the lecturer is not an economist. . . It means, that he can not provide You with any economical blabla about business. . .

Instead, we'll learn how to extract implicit, previously unknown and potentially useful information from data, i.e. **data mining**.

Why to mine data?

- There are often useful information hidden in data.
- Lots of data has been/is being collected, impossible to analyze all of it manually for managers.
- Competitive pressure, i.e. more knowledge about customers and goods, better business decisions.



The goal of the course

Provide You with basic data mining – pattern mining, forecasting, recommendation and web mining – techniques.

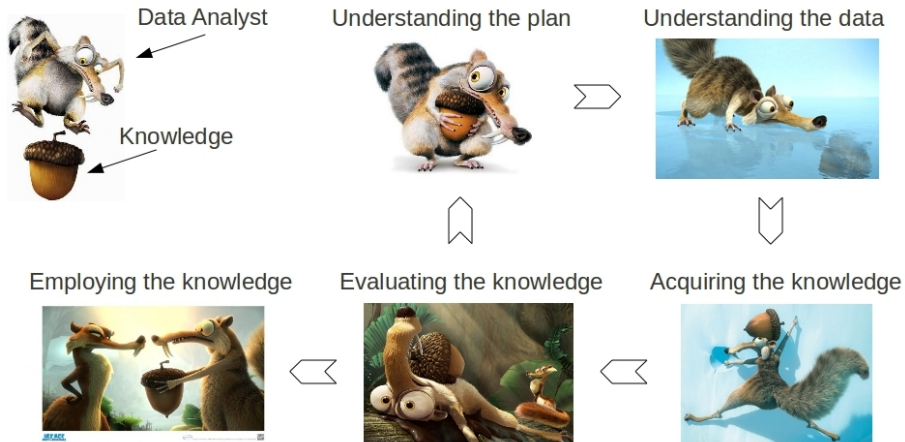
- Solid theoretical background
- Some freely available software tools
- Practical demonstration on real-world data

It may happen, that some contents discussed during the lecture can be seen/heard on other courses as well but the overlap should be small.

- Master courses at ISMLL
 - Machine Learning
 - Advanced Topics in Machine Learning
 - Bayesian Networks
- Business Intelligence (Bachelor course at ISMLL)
- Database Systems and Semantic web related courses at UHi

- ① The CRISP-DM methodology
 - The KDD process; Data pre-processing; Data mining basics.
- ② Patterns
 - Clustering; Frequent Itemset Mining; Association Rules; Sequence Mining.
- ③ Time series, Sequences
 - Time series forecasting; Time series/Sequence classification.
- ④ Recommender Systems
 - User feedback; Item vs. Rating prediction; Collaborative filtering & Content-based filtering; Context-aware recommendations.
- ⑤ Web and Social mining
 - Tag recommendation; Link prediction; Trust prediction; Opinion mining.

The CRISP-DM methodology



Patterns

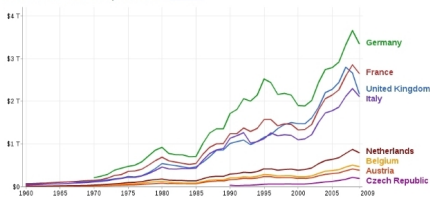
- Which items appear together in one basket frequently?
- Which customers have similar shopping behaviours?



Time series, Sequences

- Is there any trend or seasonality in data?
- How to forecast future values?
- How to classify customers based on their credit card usage?
- How to detect anomalies?

Gross Domestic Product
GDP in current U.S. dollars. Not adjusted for inflation. [More info](#)



Data source: [World Bank, World Development Indicators](#) - Last updated February 11, 2011

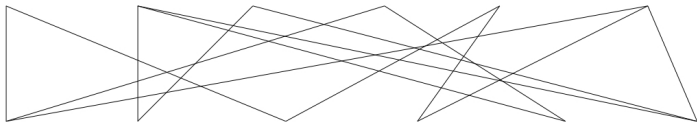
<http://www.google.com/publicdata>



<http://markets.ft.com/>

Recommender Systems

- Which “unseen“ items would the user be most likely interested in?
- How would the user rate an “unseen” item?



Web and Social Mining

The image shows two screenshots from a web browser. The top screenshot is from an Amazon product page for the Canon Powershot SD1300 IS. It features navigation tabs for 'Compare Prices', 'Read Reviews (22)', 'View Details', and 'Write a Review'. The main content area displays the product name, a star rating, and a price range of \$179.95 to \$219.95. A large, tilted watermark 'http://www.0.epinions.com' is overlaid on the page. The bottom screenshot is from a Last.fm page for the artist Shakira. It shows a 'Tags' section with various music genres and styles listed, such as 'pop', 'rock', 'latin', and 'female vocalists'. A tilted watermark 'http://www.last.fm' is overlaid on this page as well.

Compare Prices Read Reviews (22) View Details Write a Review

Showing 1-13 of 13 reviews

Sort by Product Rating

Product Rating: ★★★★★

Canon Powershot SD1300 IS
by [Howard Crech](#) (LEAD) in Electronics, (POPULAR AUTHOR)

Pros: Ultra-compact, stylish, and user friendly
Cons: No optical Viewfinder, slight tendency toward over-exp

Price Range: \$179.95 - \$219.95

Canon Powershot Digital Elph SD1300IS Digital Car Point-and-Shoot!
by [noangels](#) (POPULAR AUTHOR) - Top 1000, Oct 10 '10

Pros: Excellent photo quality, good battery life, easy to use,
Cons: No viewfinder, continuous shot speed could be faster

Price Range: \$104 - \$250

Artist
Biography
Pictures
Videos
Albums
Tracks
Events
News
Charts
Similar Artists
Tags
Listeners

Music » Shakira » Tags

Tags

00s 90s acoustic alternative amazing amazingvoice andean beautiful best best songs ever brilliant cantautor catchy colombia colombian dance dance-pop deep diva eighth wonder of the world electronic english experimental female female vocalist female vocalists female vocals fierce fun fusion genius great hip-hop hips hot latin latin pop latin rock latino legend love poet political pop pop rock pop-rock powerful rnb rock rock en espanol rock en español sexy shakira singer-songwriter spanish synth talented talented ladies world yes

People You May Know

See other people you may know below.

Ing. Ph.D. at Il-SAS
In Common: 5 shared connections

Software Analyst
In Common: 2 shared connections

Consultant at Ariba Inc
In Common: 3 shared connections

Research Assistant at Slovak University of Technology
In Common: 3 shared connections

Artist » Shakira » Tags

pop pop rock pop-rock powerful rnb rock rock en espanol rock en español sexy shakira singer-songwriter spanish synth talented talented ladies world yes

First of all: Please, **sign up** for the course in **LSF!**

- Lectures
 - provided by Tomáš Horváth
 - Every Wednesday: 10-12 ct. in B26
 - Every even¹ Thursday: 8-10 ct. in B26
- Tutorials
 - provided by Osman Akcatepe
 - Every odd² Monday: 14-16 ct. in B26
- Final mark will be computed as
 - 20% from the tutorials
 - 80% from the **final examination**

¹Even weeks of the semester, i.e. 2nd, 4th, 6th, etc.

²Odd weeks of the semester, i.e. 1st, 3rd, 5th, etc.

Thanks for Your attention!

Questions?

horvath@ismll.de

