Business Analytics - Course Introduction

Dr. Josif Grabocka

ISMLL, University of Hildesheim

Business Analytics

A digital world (vouchercloud, 2016)

- **Every day**: 2,500,000,000,000,000,000 (quintillion) Bytes of data
  - 10 million blue-ray discs, or 4 Eiffel towers
- 90% of the world's data created in last 2 years
- **Every minute:**
  - 204,000,000 million emails sent
  - 216,000 Instagram posts
  - etc...

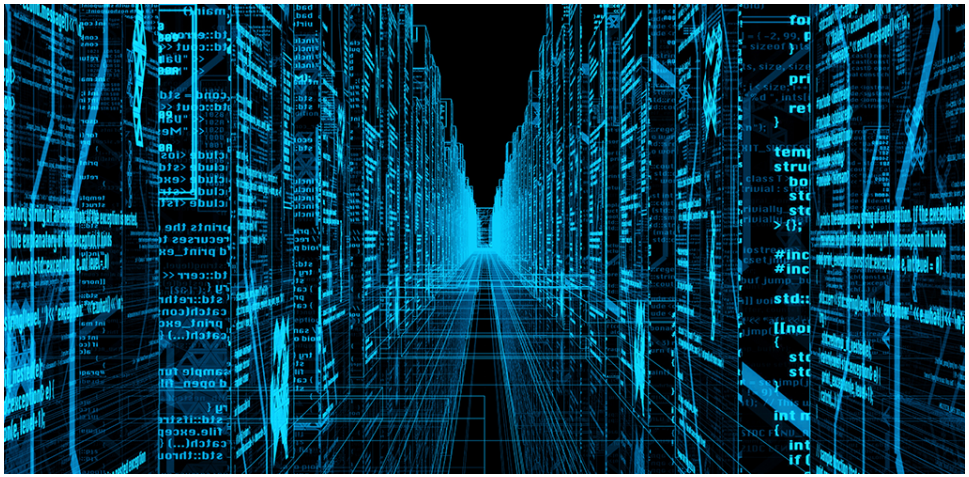


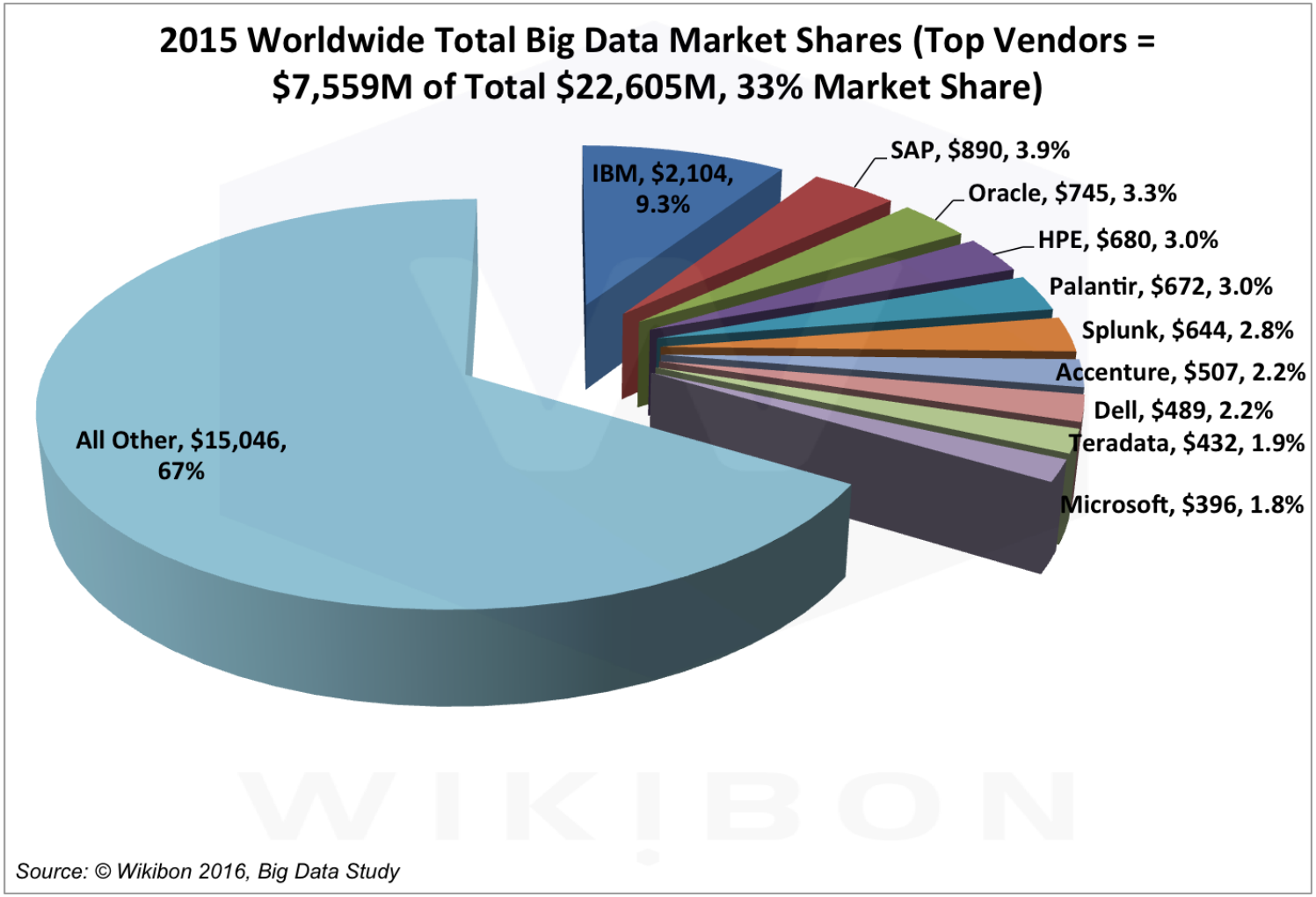
Illustration: `bigdatasp.com`

In 2018 ...

- Search:
  - 3.7 billion humans use internet
  - 5 billion searches **per day** (Google 3.5 billion, others 1.5 billion)
- Social, **every minute**:
  - Youtube users watch 4,146,600 videos
  - 2 billion Facebook users post 510,000 comments and 293,000 statuses
- Services, **every minute**:
  - Weather channel receives 18 million forecast requests
  - Uber riders take 45,788 trips

Source: Forbes, 2018

Market of Big Data



Business Analytics

- **Business Analytics** is a metaphor, referring to mining business data in order to boost revenues, through:
  - gaining insight and drive planning (**decision support systems**)
  - predicting sales, etc ... (**predictive analytics, forecasting**)
  - improving advertisements (**recommender systems**)
  - ...

Course Content

- **Time-series forecasting**: Predicting future sequentially ordered measurements
- **Predictive modeling**: The data-driven prediction tasks, state-of-the-art classification and regression models, GBDT
- **Recommender Systems**: What to recommend? To whom?
- **Learning-to-rank**: Ranking models for information retrieval systems, e.g. search engines
- **Anomaly/Outlier Detection**: Identify items/patterns that are unlikely to occur

Example 1: Time-series Forecasting

- Given the **past** four years of electricity expenditure,
- can you predict the total expenditure for the **future** months?

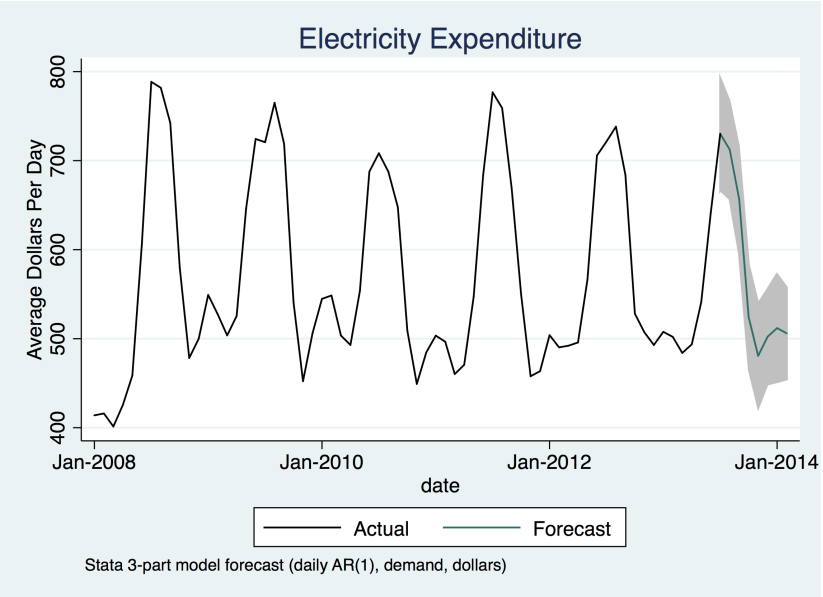


Illustration: `thinkinator.com`

Example 2: Predictive Modeling for Credit Card Fraud

- Feat1: Type of card
- Feat2: Transaction time
- Feat3: Total amount of transactions within the same day
- Feat4: Number of transactions within the same day
- Feat5: Number of failed transactions within the same day
- Feat6: Number of transactions in the last 5 days
- Target: **Fraud** or **Non-fraud**

Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Target
Visa	12:00	140	2	0	1	**Non-fraud**
Visa	23:00	2375	17	8	25	**Fraud**
Mastercard	24:00	1800	16	9	13	**Fraud**
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Visa	23:50	1676	18	8	31	**Predict it!**

Example 3: Recommender Systems

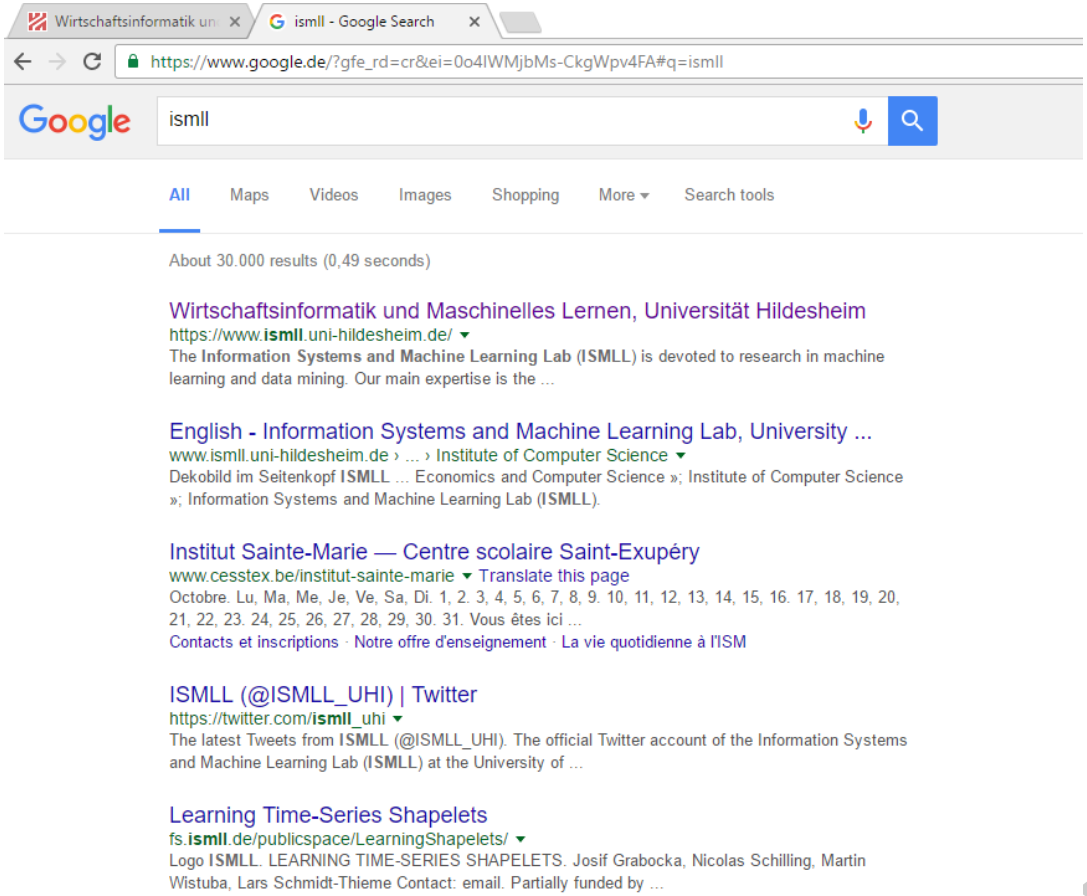
		item	1	2	3	4	5	6	7	8	9	...
		year	'95	'95	'17	'05	'15	'08	'95	'83	'16	
		action	−	+	−	+	−	−	−	−	−	
		children's	+	−	−	−	−	−	−	+	−	
		...										

	gender	age	...
user 1	m	24	
2	f	53	
3	m	23	
4	m	24	
5	f	33	
6	m	42	
⋮			

	item	1	2	3	4	5	6	7	8	9	...
user 1		4			3						
2		1									
3			4								
4						5					
5									2		
6									4		
⋮											

Example 4: Learning to Rank

- Given a list of queries and subsequent documents results
- Given the relevance of documents within queries
- Learn to rank documents of a future queries by their relevances



Example 5: Anomaly/Outlier Detection

- Detect patterns that do not follow the "usual" trend
- Items that have a low probability of occurrence
- The problem is relevant in particular for predictive maintenance, i.e. fault detection

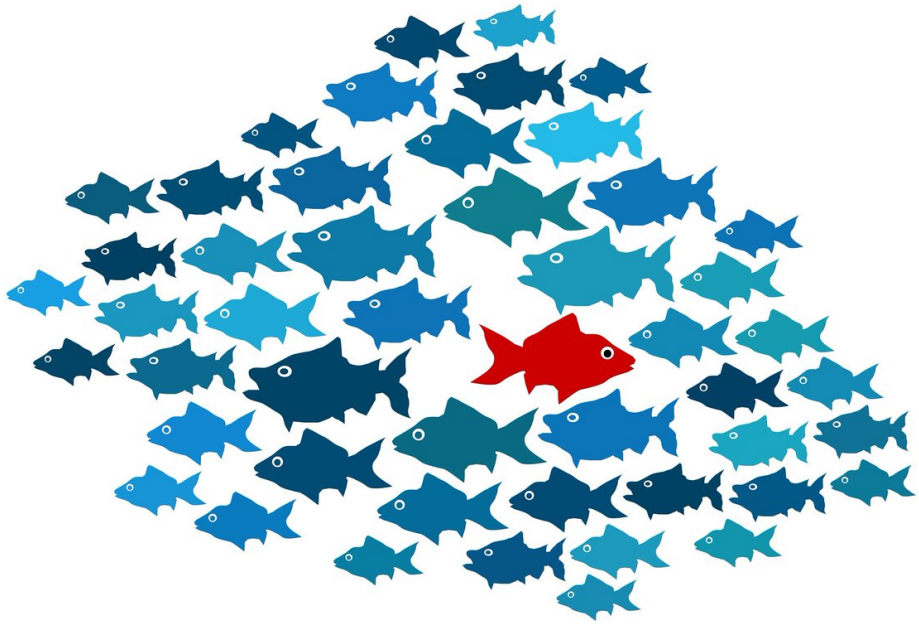


Illustration: Sergio Santoyo

Exercises and Tutorials (Rafael Drumond, C036 Spl)

- Exercises are released every Thursday afternoon (after every lecture)
- The deadline is Wednesday midnight (with a grace-time of 12 hours)
- You must submit on the LearnWeb Platform to receive your grade
- If you want to receive your sheet back with the corrections you must post on the post-box located inside Samelsomplatz building (from the entrance to the right) 54 and 55 (depending on your group)
- Exercise grades: Maximum of 4 bonus points for the exam.

Plagiarism

- When suspected of plagiarism:
  - The incident will be reported to the University's Disciplinary Board
  - The standard measure is to expel a student from the program

Illustration: `dopessays.com`

Exam and Credit Points

- There will be a written exam at end of term (ca. 2h, 4 problems).
- The course gives 6 credits
- The course can be used in
  - International Master's Program in Data Analytics.
  - Wirtschaftsinformatik MSc / Wirtschaftsinformatik / Gebiet BI
  - IMIT MSc. / Informatik / Gebiet KI & ML
  - as well as in both BSc programs.

Some Books

- Travor Hastie et al. (2004):
  The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, Springer.
- Rob J Hyndman et al. (2014):
  Forecasting: Principles and Practice, Springer.
- Charu C. Aggarwal (2016):
  Recommender Systems, The Textbook, Springer.