

Tackling Challenging Problems in Academia & Industry - An Interdisciplinary Approach

Dr. Kurt Stockinger
Associate Professor (Dozent) of Computer Science
Director of Studies in Data Science
ZHAW Datalab

SDS 2015, Winterthur, June 12, 2015

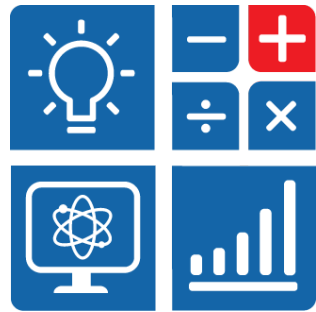
Outline

- Overview of ZHAW Datalab
- Continuing Education:
 - DAS Data Science
- Applied Research Projects:
 - Market Monitoring
 - Sentiment Analysis
 - Face Recognition
- Conclusions

- School of Engineering dates back to **Technikum** in Winterthur (founded in 1874)
- **ZHAW is a merger** of 8 universities of applied sciences in canton Zurich (in 2009)
- ZHAW is largest university of applied sciences in Switzerland with more than 11,000 students
 - Campuses in Zurich, Winterthur and Wädenswil (around Lake Zurich)
 - We do **applied research projects with industry partners**
- Most famous lecturer:
 - 1901: **Albert Einstein**



Datalab = Data Science @ ZHAW



datalab
dlab.zhaw.ch

- One of the first **Data Science Laboratories** in Europe
- Collaboration of computer scientists, statisticians, mathematicians and physicists for solving **Data Science** problems in **academia and industry**:
 - Institute of Applied Information Technology
 - Institute of Data Analysis and Process Design
 - Center for Social Law
 - Institute of Applied Mathematics and Physics (new)
 - Institute of Applied Simulation (new)
- About 60 ZHAW Datalab Associates

ZHAW Datalab Board

Operational Leadership



Dr. Thilo
Stadelmann, Head
of the board (1)



Dr. Oliver Dürr,
Deputy head of the
board (2)

- (1) Institute of Applied Information Technology
- (2) Institute of Data Analysis and Process Design
- (3) Center for Social Law
- (4) Institute of Applied Mathematics and Physics (new)
- (5) Institute of Applied Simulation (new)

Managing Board



Prof. Gerold
Baudinot (1)



Prof. Dr. Martin
Braschler (1)



Dr. Mark
Cieliebak (1)



Dr. Oliver Dürr
(2)



Prof. Dr. Rudolf
Marcel
Fuechslin (4)



Dr. Krzysztof
Kryszczuk (5)



Prof. Dr. Kurt
Pärli (3)



Prof. Dr.
Andreas
Ruckstuhl (2)



Dr. Thilo
Stadelmann (1)



Dr. Kurt
Stockinger (1)

Outline

- Overview of ZHAW Datalab
- Continuing Education:
 - DAS Data Science
- Applied Research Projects:
 - Market Monitoring
 - Sentiment Analysis
 - Face Recognition
- Conclusions



Diploma of Advanced Studies in Data Science

Content – 3 modules (part time, one afternoon + evening per week):

- **CAS Information Engineering (CAS IE):**
 - Information Retrieval & Text Analysis
 - Data Warehousing & Big Data
- **CAS Data Analysis (CAS DA):**
 - Statistical Inference, Regression, Time Series Analysis
 - Descriptive Statistics, Clustering, Classification
- **CAS Data Science Applications (CAS DASA):**
 - Machine Learning
 - Data Visualization
 - Data Protection Laws and Data Security
- More information (course description and information event):
 - <http://www.engineering.zhaw.ch/nc/de/engineering/weiterbildung/programmdetails.html?i=T903680&qu=0>
 - <http://www.zhaw.ch/de/engineering/weiterbildung/infoabend-weiterbildung.html>



Parallel Model, ~ 1 year

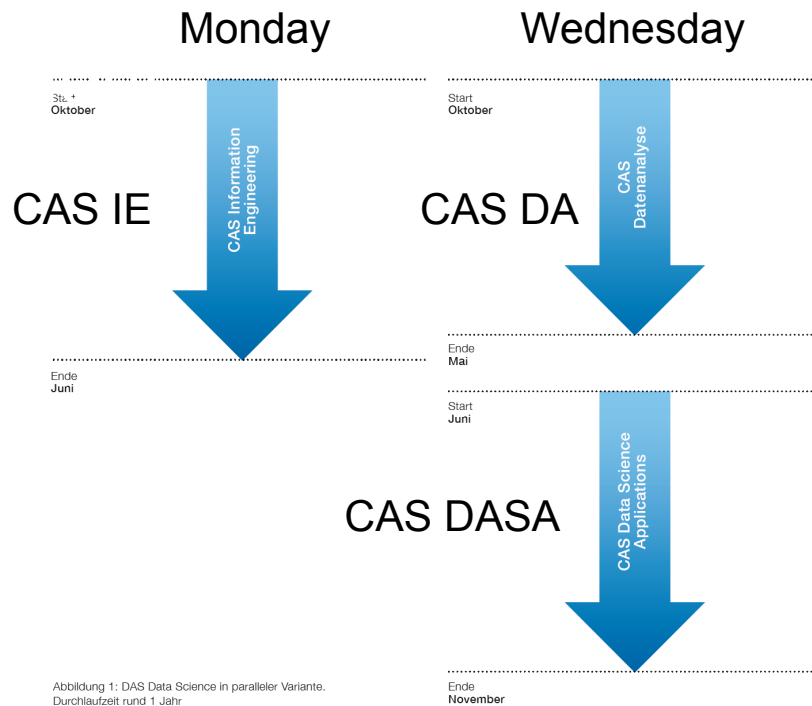


Abbildung 1: DAS Data Science in paralleler Variante. Durchlaufzeit rund 1 Jahr

Consecutive Model, ~ 2 years

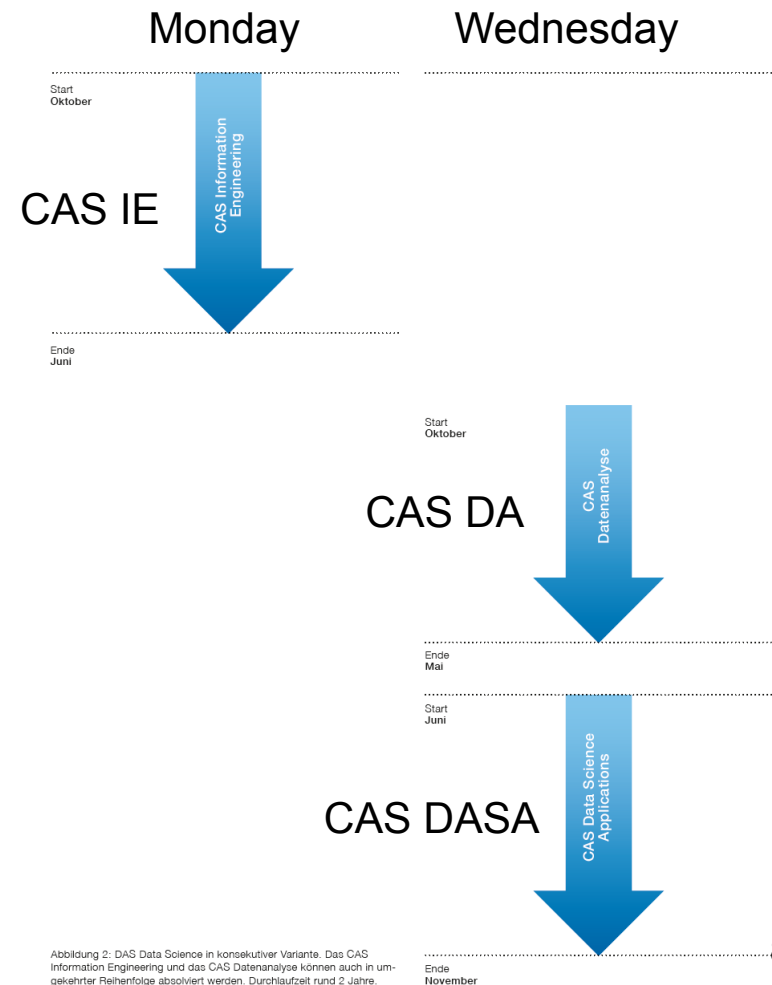


Abbildung 2: DAS Data Science in konsekutiver Variante. Das CAS Information Engineering und das CAS Datenanalyse können auch in umgekehrter Reihenfolge absolviert werden. Durchlaufzeit rund 2 Jahre.

Outline

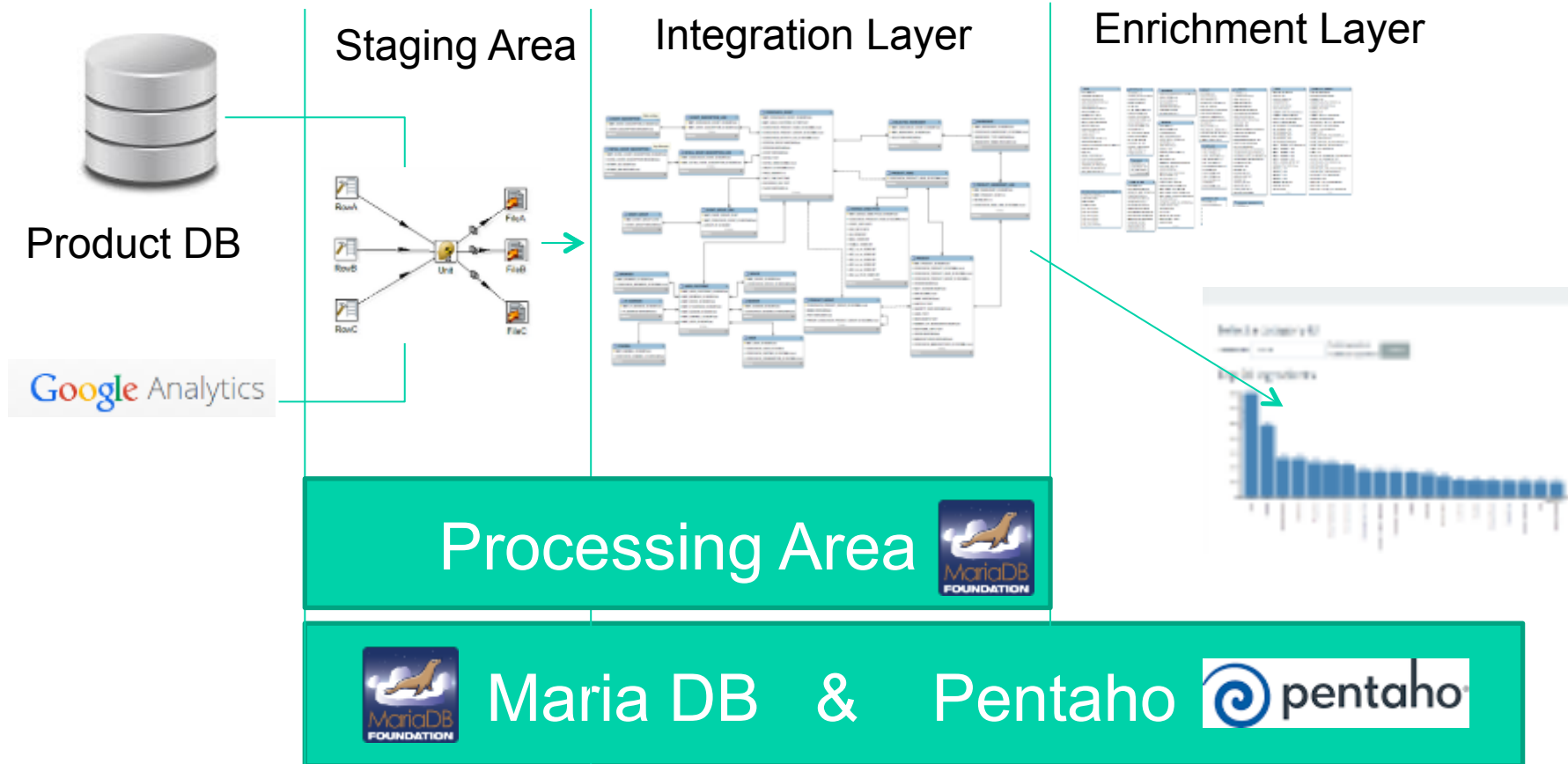
- Overview of ZHAW Datalab
- Continuing Education:
 - DAS Data Science
- Applied Research Projects:
 - Market Monitoring
 - Sentiment Analysis
 - Face Recognition
- Conclusions

Project with Start-Up: Market Monitoring

- Consumer information about products with focus on **groceries** und **cosmetics**
- Product portal with about **tens of million users** and **tens of millions products**
- Users are **anonymous**
- **No buying** of products required
- Goal:
 - Customers can get information about **healthy products** with **sustainable production**



Data Warehouse Architecture

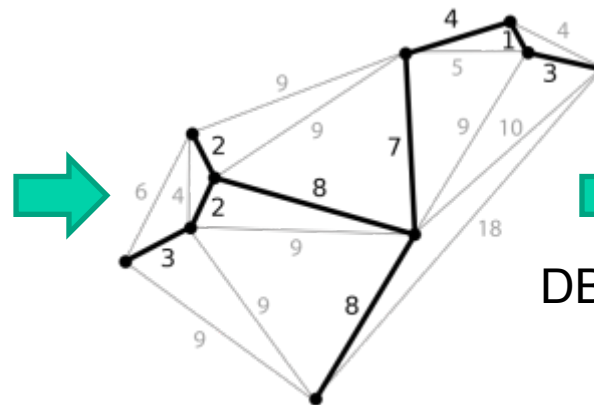


Clustering of Products based on Clickpaths

Clickpath Logs

CLICKPATH_ID	PRODUCT_HEAD_ID	PRODUCT_NAME
238	13724517	CATRICE Re-Touch Light-Reflecting Concealer - 001 Porcelain
238	267716	Iskantan Soft Compact Powder Natural Look Farbe 0 Transparent
285	12898275	MAYBELLINE Jade Kompaktpuder »Supersley 24H Puder 0 px
284	763928	Fine Loose Mineral Powder
284	13139523	Annamarie Börlind - Spezialpflege Ölbasis Tagespflege
284	245287	Dr. Hauschka Translucent Face Powder loose
284	886633	Hildegard Brauhmann - Jeunesse Puder Make up mittel
284	518089	Artdeco Make-up Gesicht Perfect Tint Illuminator: 0 1 08
238	816675	perfect face anti-red base
297	959862	Maybelline Jade Supersley 24h
287	729445364	Alverde - Mineral Make up 01 Naturelle
357	764883530	Missha BB Bomber
357	12668829	BRTC Jasmine Water BB Cream
357	1182905	skin Factor Make-Up Gesicht Ageless Elvir 2in1 FoundationNr. 60 Sand 1 08
338	372226	Demossence cream cover
357	818985	The Oriental Oild BB Cream Plus
357	818985	The Oriental Oild BB Cream Plus
357	13913999	BB Cream Judith Williams
357	826382	Skin79 Lowly Oil BB Cream
357	13625233	Avon Ideal Flawless BB Cream
357	10511238	Skin79 super* BBish Balm Vital Orange
357	10462262	Clinique - Age Defense BB Cream SPF 30 Color 01

Product Distances



Product Cluster



$$d(A, B) = \frac{\#(A \cup B) \in \text{Clickpaths}}{\# A \in \text{Clickpaths} + \# B \in \text{Clickpaths}} \cdot 2$$



- Lacura Beauty Make up
- Lacura Beauty – Compact Powder
- Lacura Beauty Feuchtigkeits-Make-up Lacura Beauty
- LACURA BEAUTY Make-Up
- Lacura Beautycare Make up beige
- Lacura
- Lacura Beauty
- Lacura Compact Powder

- Mary kay cc cream
- mary kay full coverage foundation
- Mary Kay Foundation Primer
- Mary Kay Concealer
- Mary Kay Luminous-Wear Liquid Foundation
- Mary Kay Translucent Loose Powder
- MARY KAY® MINERAL POWDER FOUNDATION
- Mary Kay Matte-Wear Liquid Foundation

- La Prairie Make-up Foundation Powder Cellular Treatment Foundation Powder FinishIvoire 1 Stk.
- Sisley - phyto teint eclat
- sisley Phyto Teint Perfect Compact Lissant
- La Prairie Make-up Foundation Powder Anti-Aging Foundation SPF 15Shade 200 30 ml
- La Prairie Skin Caviar Concealer Foundation SPF 15
- Sisley Make-up Teint Phyto Touche Or 50 ml
- La Prairie - Light Fantastic Cellular Concealing

- Manhattan - 2 in 1 Perfect Teint Powder & Make up, Sunbeige 21
- 2in1 Perfect Teint Powder & Make Up
- Manhattan Soft Compact Powder

Outline

- Overview of ZHAW Datalab
- Continuing Education:
 - DAS Data Science
- Applied Research Projects:
 - Market Monitoring
 - Sentiment Analysis
 - Face Recognition
- Conclusions

Contact: Dr. Mark Cieliebak



What is Text Analytics?



Sentiment Analysis



Named Entity Extraction



Spelling Correction

**Goal: Turn text
into information**



Text Summarization

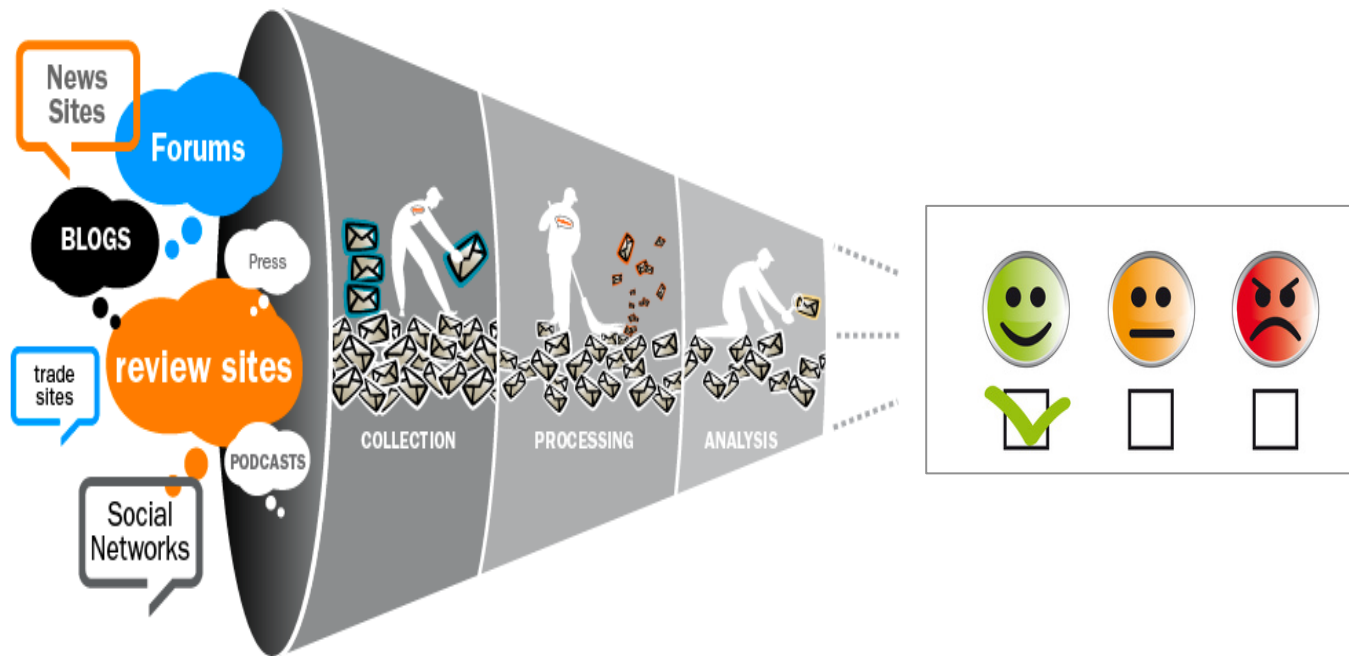


Machine Translation



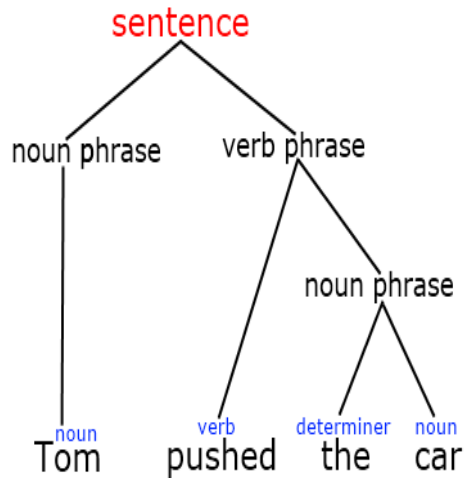
Q&A

Sentiment Analysis

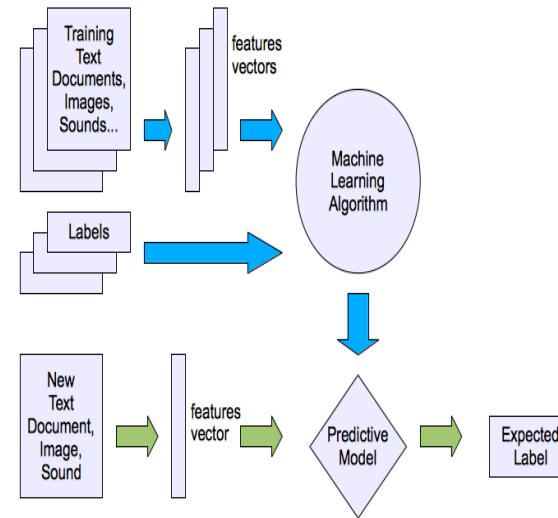


Approaches to Sentiment Detection

Rule-Based



Corpus-Based



Rule-based:

- Find sentence structure, grammar etc. to construct rules to determine sentiment [9]
- Works well but lots of manual effort for each language
- Impossible for Twitter with informal language and slang

Corpus-based: Use training data and apply machine learning algorithms [10]

Project Results

- Evaluation of **commercial and scientific tools** for sentiment analysis:
 - They reach **F1-score** (combination of precision and recall) of **60%** on arbitrary documents
 - **70%** on tweets
- Implemented **sentiment system**:
 - 8th at both SemEval 2014 and 2015
 - About 50 international participants
 - 2014 **F1-Score**: Winner: **70.94%**, ZHAW: **67.54%**
 - 2015 **F1-Score**: Winner: **64.84%**, ZHAW: **61.62%** (different data set)
- Developed a **generic framework for text classification**:
 - Spam detection, sentiment, fraud detection etc.
 - Applied for sentiment analysis in various languages



Live Demo: SentiZHAW

The screenshot displays the SentiZHAW live demo interface. At the top left, there is a profile for @SentiZHAW with a blue robot-like avatar and the text "Für Hilfe: „@SentiZHAW HELP“". In the center, the topic "Apple" is shown with a remaining time of "00:49min". Below this, a tweet from Dracos Lemming (@LemmingKGTM) is featured, which is a retweet of @Mark_Beech about iPhone sales. A green plus icon and the text "Der Computer meint:" are next to the tweet. At the bottom, a bar shows "Positive Tweets 81" and "214 Negative Tweets". Below this bar, a "Live Feed | Analyzierte Tweets: 295" section shows three tweet snippets:

- mobile followers** @mobilefollower3: RT @beretainhawk: Oh, wow! It seems I'm the best dressed at the party! #SUPERSTARLIFE <https://t.co/4VVlg6WQh6> <http://t.co/1tqPDC4HKm>
- ninifitit** @ninfifit: YAAASSSS!!! It's FREE NOW -3! <http://t.co/p17SUwJ325> #LagName
- محمد زهنيا** @UsenchukZhenya: RT @Tmood55: Even if I knew that tomorrow the world would go to pieces, I would still plant my apple tree. Martin Luther

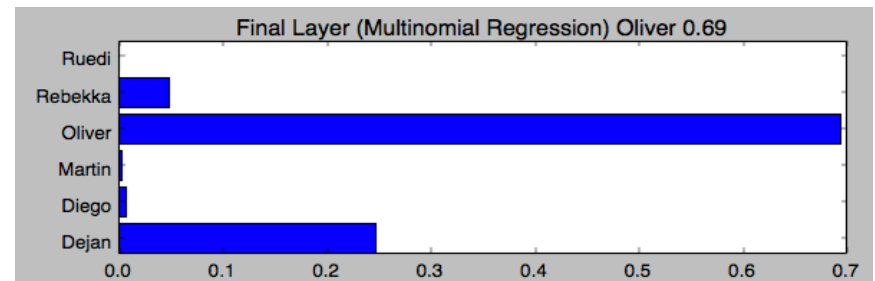
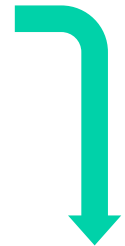
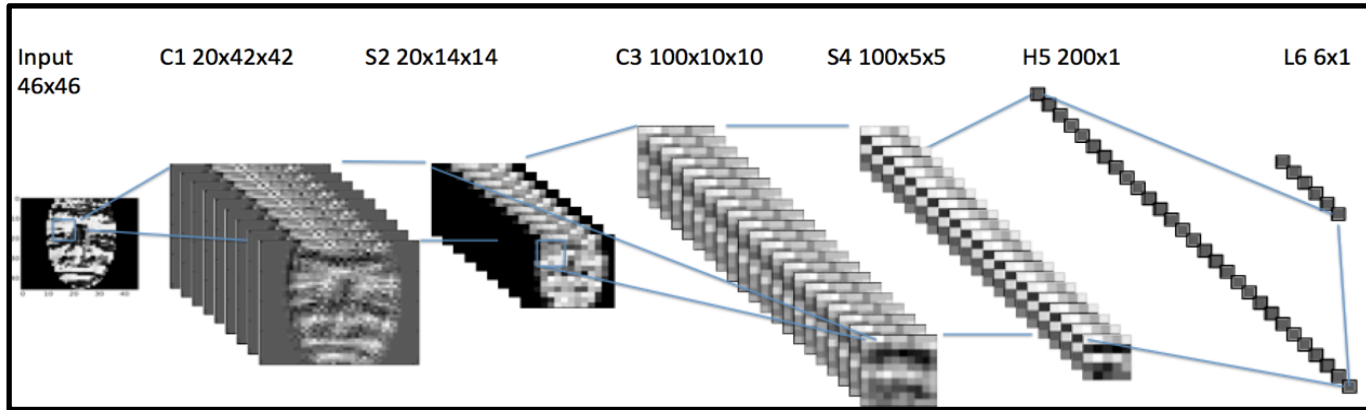
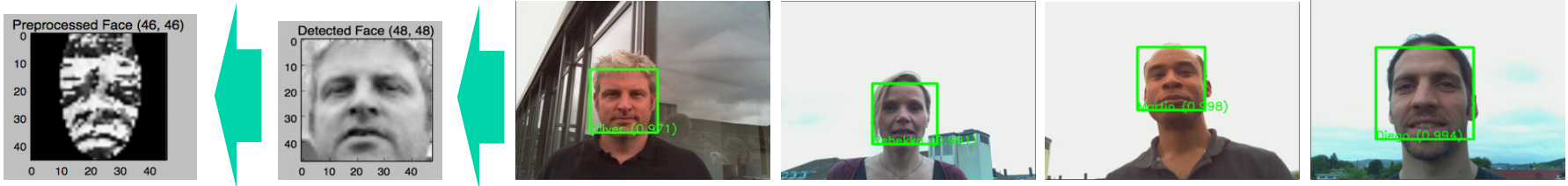
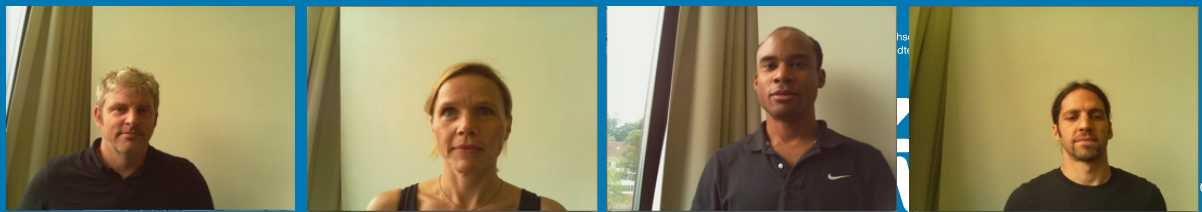
- Overview of ZHAW Datalab
- Continuing Education:
 - DAS Data Science
- Applied Research Projects:
 - Market Monitoring
 - Sentiment Analysis
 - Face Recognition

Contact: Dr. Oliver Dürr

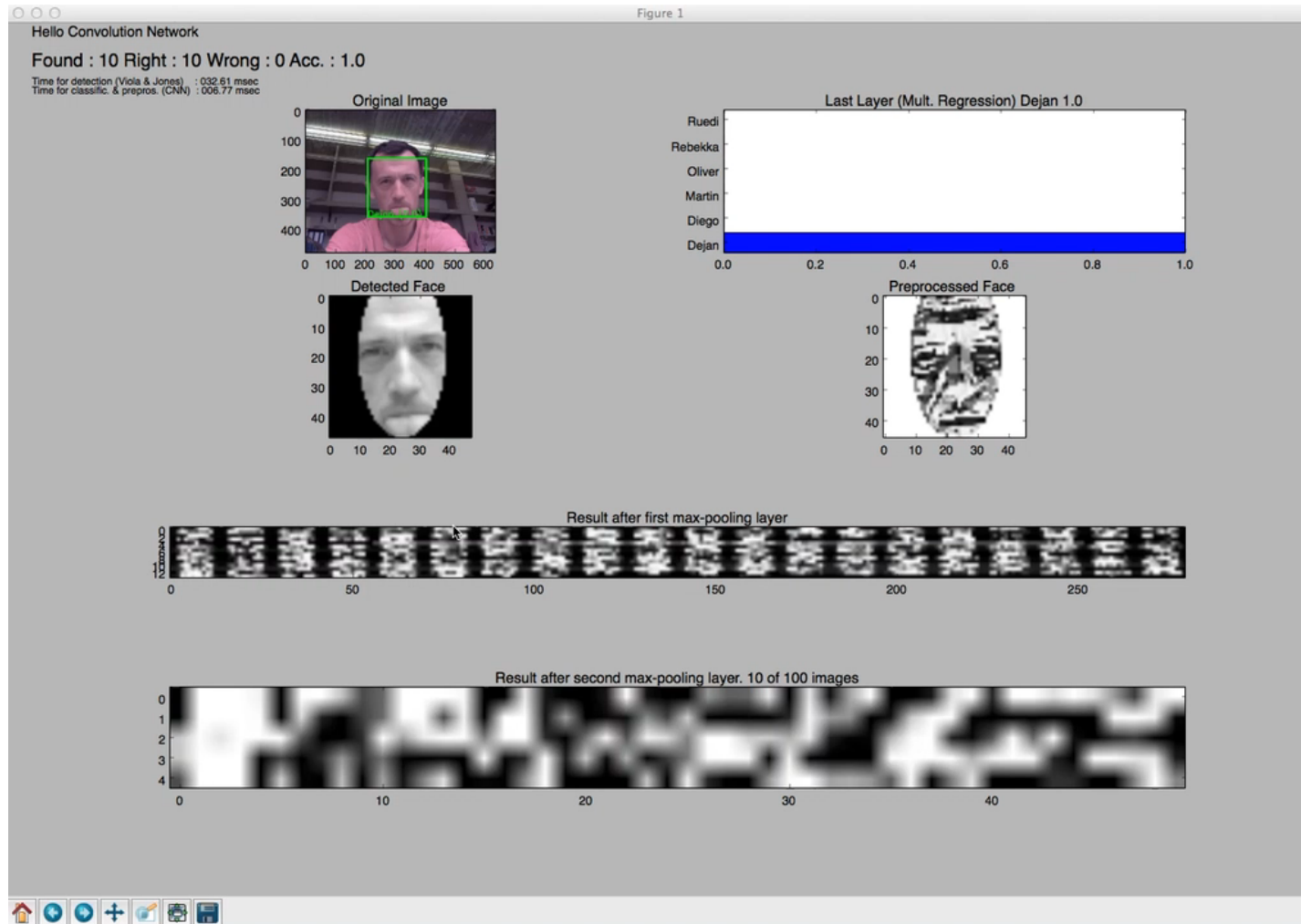


- Conclusions

Deep Learning on Raspberry Pi for Face Recognition



The Neural Network at Work on New Data



- Deep Learning works on a Raspberry Pi
- Better performance than standard openCV pipeline (open source library for computer vision)
 - Speed of classification 105 msec vs 535 msec
 - Accuracy 99.5% vs 96.8%

Method	Accuracy	Classification Time [msec]	Enrollment Rate N_e/N	Total Time Per Face [msec]
CNN ($p_0=0.85$)	99.59%	105 +/- 8	250/278	529 +/- 64
CNN ($p_0=0.00$)	97.48%	105 +/- 8	278/278	529 +/- 64
Fisherfaces (no al.)	88.5%	54 +/- 11	278/278	511 +/- 89
Fisherfaces (al.)	96.87%	535 +/- 89	192/278	1006 +/- 18

Accuracy and performance for various approaches.

Conclusions

- **Lessons learned** in large-scale data analysis:
 - Start with explorative analysis for early insights
 - Build a data warehouse to integrate and cleanse core data
 - Iteratively refine and update data marts for fast query processing
 - Apply machine learning / deep learning
 - Performance optimization of queries and machine learning takes time
- **Data Science projects** require teams with very diverse skills
 - **ZHAW Datalab**:
 - Enables **multi-disciplinary approach** of computer science, statistics, etc.
 - Applied research projects are very **attractive for industry partners**
 - **Educates Data Scientists** of the future

Appendix

Opportunities for Collaboration with ZHAW Datalab

- **EU-Projects:**
 - Funding via European Union
 - „Calls“ for specific topics with fixed deadlines
 - E.g. Big Data Research
- **CTI-Projects** (Commission for Technology and Innovation):
 - Funding of applied research projects through Federal Government
 - Collaboration between industry partner and university
 - No specific topics
 - 10 submission deadlines per year, feedback within 2-3 weeks
- **Direct Funding:**
 - Industry partner funds ZHAW Datalab directly