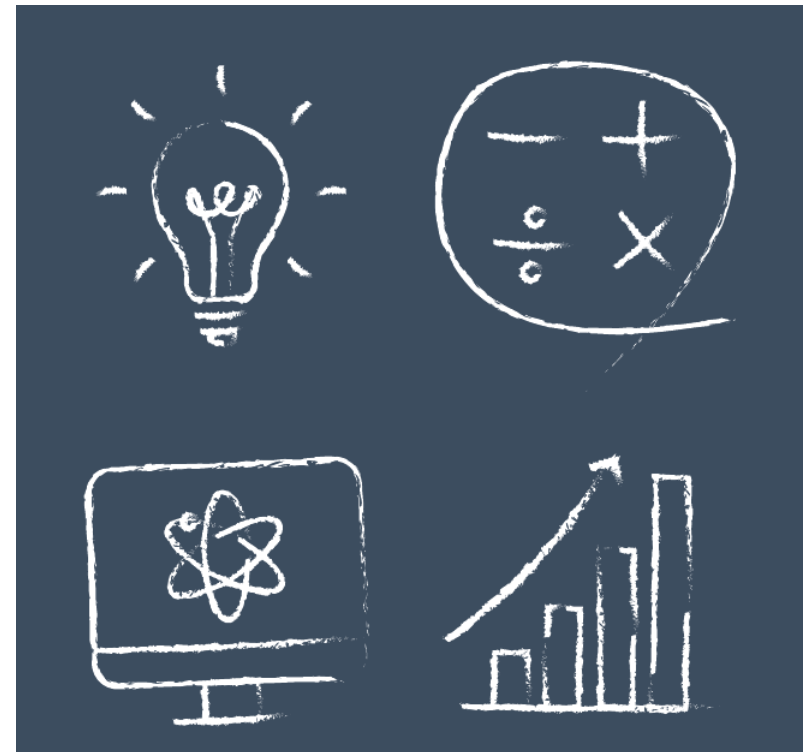


# Datalab – The ZHAW Data Science Laboratory

**Dr. Thoralf Mildenerger**

Institut für Datenanalyse und Prozessdesign  
Zürcher Hochschule für Angewandte Wissenschaften

Data Science in Studium und Lehre 2018  
HIS-Institut für Hochschulentwicklung  
19. 9. 2018





# Übersicht

- Die Zürcher Hochschule für angewandte Wissenschaften
- Das Datalab – Organisation und Themen
- Data Science in Forschung und Entwicklung
- Aktivitäten in der Lehre
- Berufsbegleitende Weiterbildung

# Die Zürcher Hochschule für angewandte Wissenschaften (ZHAW)

- Die **Zürcher Hochschule für angewandte Wissenschaften (ZHAW)** ist eine der grössten Fachhochschulen der Schweiz
- Zusammen mit der Zürcher Hochschule der Künste, Pädagogischen Hochschule Zürich und Zürcher Hochschule für Wirtschaft Teil der **Zürcher Fachhochschule (ZFH)**
- **Ca. 13'200 Studierende**
- 3 Standorte: Winterthur, Zürich, Wädenswil
- 8 Departemente:
  - Architektur, Gestaltung und Bauingenieurwesen
  - Gesundheit
  - Angewandte Linguistik
  - Life Sciences and Facility Management
  - Angewandte Psychologie
  - Soziale Arbeit
  - School of Engineering
  - School of Management and Law



# Die Zürcher Hochschule für angewandte Wissenschaften (ZHAW)

- 28 **Bachelor-** und 18 **Masterstudiengänge**
- Viele Studiengänge auch als Teilzeitstudium möglich
- Ca. 50 **Weiterbildungs-**Masterstudiengänge (MAS/EMBA/MBA), ca. 240 Diplom- und Zertifikatslehrgänge (DAS/CAS) sowie zahlreiche Weiterbildungskurse
- **Forschung und Entwicklung / Dienstleistung:** jährlich mehrere hundert Projekte mit einem Gesamtvolumen von rund 113 Millionen CHF
- verschiedene Formen der Zusammenarbeit:
  - Projekt- und Abschlussarbeiten
  - Forschungsaufträge und Dienstleistungen
  - Forschungs- und Entwicklungsprojekten

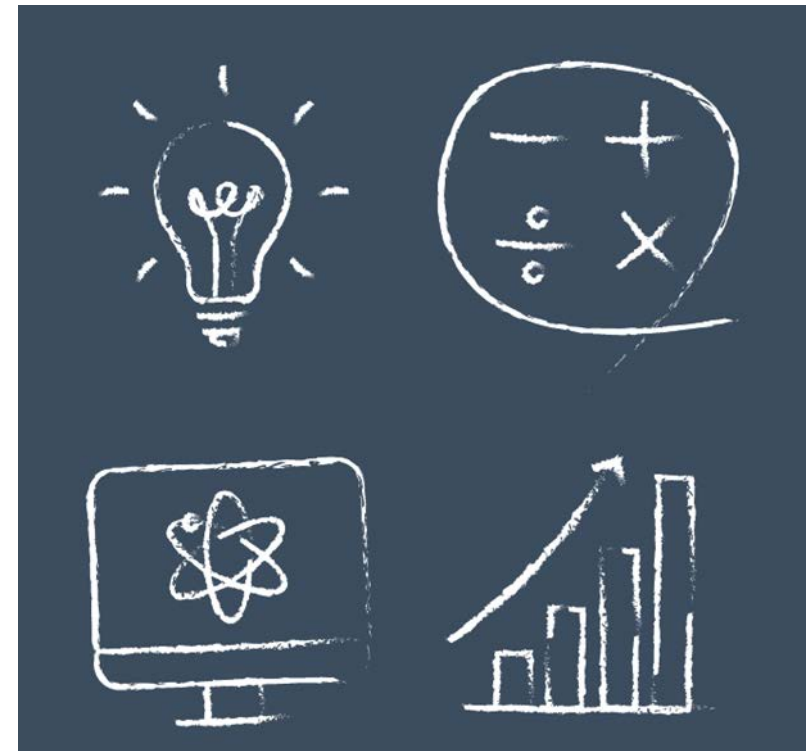
## Finanzierung:

- Beiträge des Kantons pro Studierendem/r
- Einnahmen aus Weiterbildung
- Drittmittel (F&E, Dienstleistung)
- Beiträge des Kantons proportional zu eingeworbenen Drittmitteln



# ZHAW Datalab

- Das **Datalab** ist eine **virtuelle Organization** an der ZHAW
- Mitglieder sind **Institute / Zentren**, nicht Einzelpersonen
- Das Datalab bündelt Aktivitäten im Bereich **Data Science** von verschiedenen ForscherInnen an den beteiligten Instituten
- Gegründet **2013** vom InIT (Institut für Angewandte Informationstechnologie) und IDP (Institut für Datenanalyse und Prozessdesign), beide School of Engineering.
- Mitglieder sind aktuell **5 Institute** an **3 Departementen** sowie 8 assoziierte Institute
- **Leadership Board**: 9 Personen aus den beteiligten Instituten (dies sind nicht zwingend die Instituts- oder Schwerpunktleiter)
- Derzeit 66 **Associates** auf Webseite gelistet
- Aktivitäten in **F&E, Lehre** und **Weiterbildung**



# ZHAW Datalab: Infrastruktur

- Das Datalab wird getragen von den beteiligten Instituten, die Ressourcen (geringer Finanzbeitrag, vor allem Arbeitszeit) einbringen
- Es existieren keine eigenen Räumlichkeiten
- Kein zusätzliches Budget

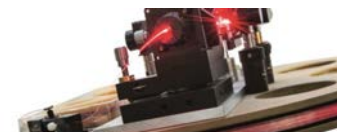
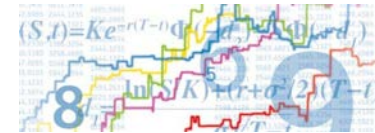
## Hardware:

- Datalab-Cluster (5 Nodes)
- Kontingent im ICC Cloudlab Cluster (7 Nodes)
- 2 Workstations mit je 2 GPUs für Deep Learning
- Zugriff auf weitere Rechenkapazitäten wie HPC-Cluster in Wädenswil



# Beteiligte Institute

- **IDP**  
Institut für Datenanalyse und Prozessdesign  
School of Engineering
- **InIT**  
Insitut für Angewandte Informationstechnologie  
School of Engineering
- **ZSR**  
Zentrum für Sozialrecht  
School of Management and Law
- **IAS**  
Institut für Angewandte Simulation,  
Dept. Life Sciences and Facility Management
- **IAMP**  
Institut für Angewandte Mathematik und Physik  
School of Engineering





# Assoziierte Institute

## **School of Engineering:**

- Institute of Embedded Systems
- Institute for Signal Processing and Wireless Communications
- Zentrum für Aviatik

## **School of Management and Law:**

- Institut für Wirtschaftsinformatik
- Institut für Financial Management

## **Departement Life Sciences and Facility Management:**

- Institut für Umwelt und Natürliche Ressourcen

## **Departement Soziale Arbeit:**

- Institut für Delinquenz und Kriminalprävention

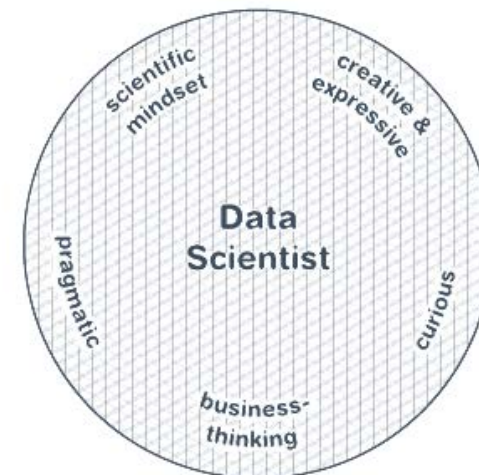
## **Departement Angewandte Psychologie:**

- Psychologisches Institut



## What is Data Science?

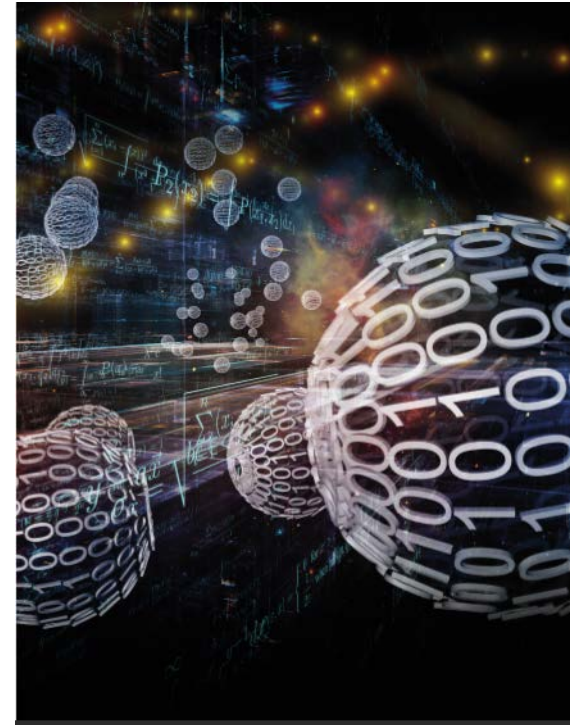
We define data science as the “unique blend of skills from analytics, engineering and communication aiming at generating value from the data itself [...]” \*



\* Stadelmann, Stockinger, Braschler, Cieliebak, Baudinot, Dürr and Ruckstuhl (2013). Applied Data Science in Europe. ECSS 2013, Amsterdam

# ZHAW Datalab: Themenfelder

- Datenbanken und Big-Data-Technologie
- Data Mining, Statistik und Predictive Modelling
- Machine Learning, besonders Deep Learning
- Graph Analytics und Network Science
- Information Retrieval und Natural Language Processing
- Simulation und Signalverarbeitung
- Business Intelligence und Visual Analytics
- Data Warehousing und Decision Support
- Kommunikation und Visualisierung von Ergebnissen
- Datenschutz, Sicherheit und Ethik
- Entrepreneurship und Data Product Design





# Forschung und Entwicklung

- Forschung und Entwicklung vor allem im Rahmen von **angewandten Projekten mit Industriepartnern**
- Finanzierung vor allem direkt durch **Industriepartner** oder durch **Innosuisse** (*Schweizerische Agentur für Innovationsförderung, früher Kommission für Technologie und Innovation*)
- “Klassische” Forschungsdrittmittel (Schweizer Nationalfonds, Horizon 2020) spielen eine geringere Rolle
- Partner sind oft **KMU**, aber auch grosse Konzerne oder Behörden
- **Projektvolumen** von ca. 15'000 CHF bis mehrere Hunderttausend CHF
- Gesamtvolumen ca. **2.5 Millionen CHF pro Jahr**
- **Publikation** von Ergebnissen in **Fachjournalen** und auf **Tagungen** (wegen Geheimhaltungsvereinbarungen nicht immer möglich).



# Forschung und Entwicklung

## Projektbeispiele:

- **Bio-SODA: Enabling Complex, Semantic Queries to Bioinformatics Databases through Intuitive Searching over Data**  
Ziel: Entwicklung semantischer Abfragen für grosse, dezentrale und heterogene Datenbanken über ein intuitives Interface  
Drittmittelgeber: SNF  
Projektpartner: Universität Lausanne, SIB Swiss Institute of Bioinformatics
- **SAVE - Smart Alarms & Verified Events**  
Ziel: Reduktion von Fehlalarmen mithilfe von Big-Data-Technologie und Machine Learning  
Drittmittelgeber: KTI  
Projektpartner: Sitasys



# Forschung und Entwicklung

## Projektbeispiele:

- **Modellbasierter Zweistufenklassifikator für Schwebestaub**  
Ziel: Klassifikation von Partikeln aus Luftproben  
Drittmittelgeber: KTI  
Projektpartner: Particle Vision GmbH
- **Stiftungsregister SR\_2**  
Ziel: Visualisierung und Suche für die schweizerische Stiftungslandschaft  
Drittmittelgeber: KTI  
Projektpartner: Univ. Freiburg, Institut VMI, NonproCons
- **FarmAI – Künstliche Intelligenz für den Farming Simulator**  
Ziel: Entwicklung eines intelligenten Gegners für ein Computerspiel mithilfe von Reinforcement Learning  
Drittmittelgeber: KTI  
Projektpartner: Giants Software GmbH

# Forschung und Entwicklung

## Projektbeispiele:

- **Deep Learning zur Erkennung von neuartigen Zelltypen**  
Ziel: Klassifikation von Zelltypen inkl. Erkennung bisher unbekannter Typen  
Drittmittelgeber: KTI  
Projektpartner: GeneData AG
- **Cleverpendeln**  
Ziel: Pendlern werden mittels App wagenscharfe Informationen über das erwartete Platzangebot in Zügen geliefert  
Drittmittelgeber: KTI  
Projektpartner: one2seven gmbh
- **Decision Support System For Predictive Maintenance of Laser Cutting Machines**  
Ziel: Vorausschauende Instandhaltung für Laserschneidemaschinen  
Drittmittelgeber: KTI  
Projektpartner: Bystronic Laser AG



# Lehre: Bachelorstudiengänge

Datalab-Associates unterrichten in zahlreichen verschiedenen **Bachelorstudiengängen**, z.B.:

- Informatik
- Wirtschaftsingenieurwesen
- Aviatik
- Verkehrssysteme
- Biotechnologie
- Umweltingenieurwesen
- Facility Management
- ...

Dadurch fließen Data-Science-Themen verstärkt in den Unterricht ein, z.B. Umgang mit grösseren und komplexen Datensätzen in Statistik-Praktika oder in Abschlussarbeiten.

# Lehre: Master

Der **Master of Science and Engineering (MSE)** ist ein gemeinsames Masterprogramm der sieben Schweizer Fachhochschulen.

<p><b>BUSINESS ENGINEERING AND PRODUCTION (BEP)</b></p> <ul style="list-style-type: none"><li>● Business Process Management</li><li>● Operation Management and Analytics</li><li>● Supply Network Management</li><li>● Product and Service Innovation</li></ul>	<p><b>ENERGY AND ENVIRONMENT (EE)</b></p> <ul style="list-style-type: none"><li>● Umwelttechnik</li><li>● Energietechnik</li><li>● Verfahrenstechnik</li></ul>	<p><b>INDUSTRIAL TECHNOLOGIES (INT)</b></p> <ul style="list-style-type: none"><li>● Produktentwicklung und Produktionstechnologie</li><li>● Werkstofftechnologie</li><li>● Mechatronik und Automation</li><li>● Embedded Systems und Mikroelektronik</li></ul>	<p><b>INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT)</b></p> <ul style="list-style-type: none"><li>● Kommunikations- und Informationssysteme</li><li>● Software Engineering und -Technologie</li><li>● Daten- und Informationsmanagement</li></ul>
<p><b>CIVIL ENGINEERING &amp; BUILDING TECHNOLOGY (CEBT)</b></p> <ul style="list-style-type: none"><li>● Nachhaltiges Bauen, Umgang mit bestehender Bausubstanz und Gebäudetechnik</li><li>● Bau- und Fertigungstechnik</li><li>● Geotechnik und Naturereignisse</li><li>● Tragwerksentwurf und Konstruktion</li></ul>	<p><b>GEOMATICS (GEO)</b></p> <ul style="list-style-type: none"><li>● Geoinformationstechnologie</li></ul>	<p><b>SPATIAL DEVELOPMENT &amp; LANDSCAPE ARCHITECTURE (SDLA)</b></p> <ul style="list-style-type: none"><li>● Raumentwicklung</li><li>● Landschaftsarchitektur</li></ul>	

**MSE**

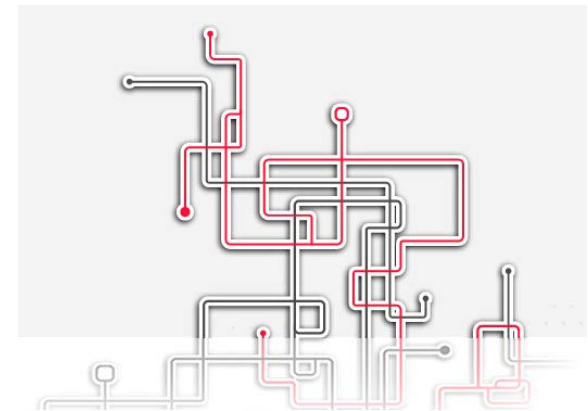
MASTER OF SCIENCE  
IN ENGINEERING



# Lehre: Master

Neue Vertiefung “Data Science” im **MSE** (Information and Communication Technologies) mit vielen einschlägigen Modulen:

- Advanced Statistical Data Analysis
- Stochastic Modelling
- Optimization
- Machine Learning
- Deep Learning
- Analysis of Text Data
- Analysis of Sequential Data
- Machine Learning in Computer Vision
- Predictive Modelling
- Kontextmodul Privatsphäre und Recht



**MSE**

MASTER OF SCIENCE  
IN ENGINEERING



# Promotionsprogramme

## PhD Network in Data Science

- Gemeinsames Doktoratsprogramm der FHs ZHAW und SUPSI mit den Universitäten Zürich und Neuchatel
- Offen für AbsolventInnen mit FH-Master
- HauptbetreuerIn an Universität, Co-BetreuerIn an FH
- Arbeit an angewandten Projekten an FH
- Dokortitel wird von Universität verliehen
- Zur Zeit ca. 10 DoktorandInnen an der ZHAW
  
- Weitere Kooperation: Università Ca'Foscari, Venedig

# Weiterbildung

Unsere Weiterbildungsangebote im Bereich **Data Science**:

## **Master of Advanced Studies (MAS) (Masterstufe)**

- **MAS Data Science (60 ECTS)**

## **Diploma of Advanced Studies (DAS) (Diplomstufe)**

- **DAS Data Science (36 ECTS)**

## **Certificate of Advanced Studies (CAS) (Zertifikatsstufe)**

- **CAS Information Engineering (12 ECTS)**
- **CAS Datenanalyse (12 ECTS)**
- **CAS Machine Intelligence (12 ECTS)**
- **CAS Statistical Modelling (12 ECTS)**
- **CAS Data Product Design (12 ECTS)**



# Weiterbildung

**Zielgruppe:** Teilnehmer mit Hochschulabschluss (auch FH) und Berufserfahrung.

**Zulassung:** Teilnehmer ohne Hochschulabschluss aber mit einschlägiger Berufspraxis können “sur dossier” zugelassen werden.

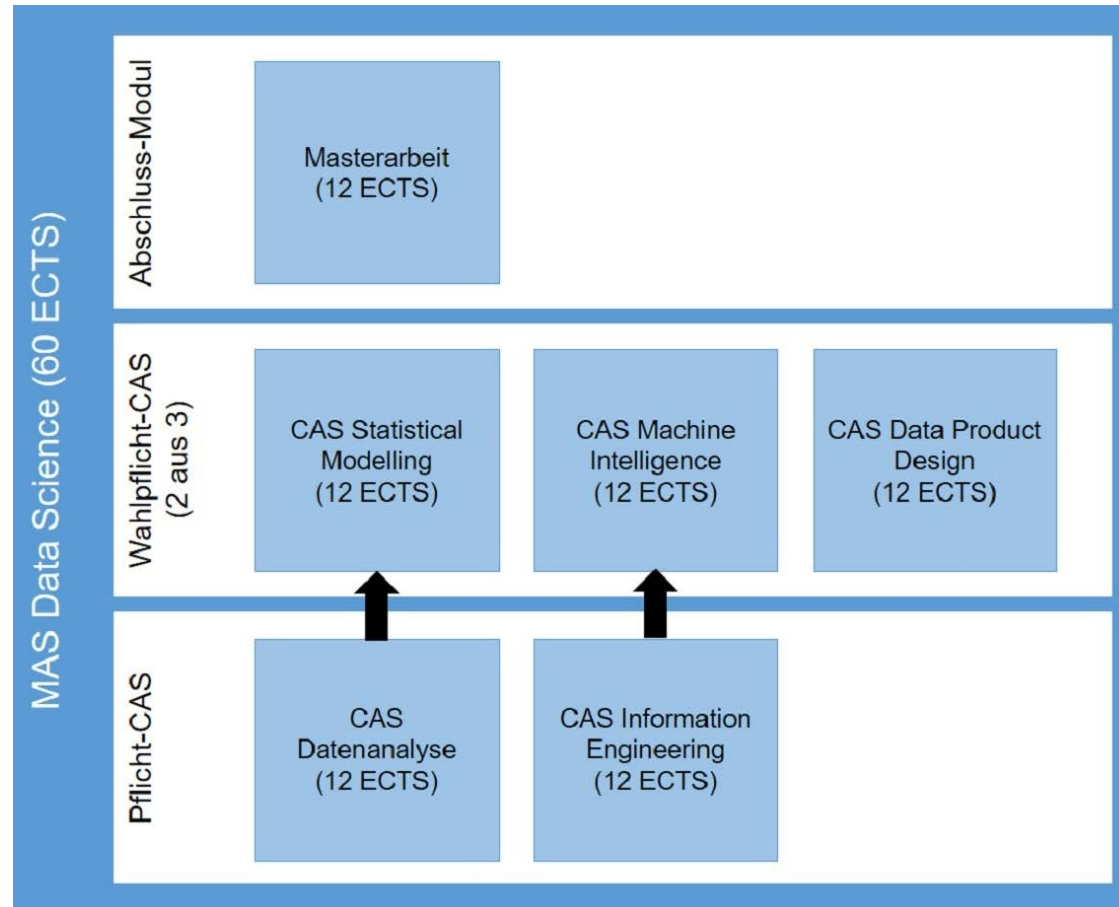
**Dauer eines CAS:** ca. 5 Monate, 1 Tag pro Woche mit 8 Lektionen à 45 min.

**Modularer Aufbau:** DAS besteht aus 3 CAS, MAS aus 4 CAS + Masterarbeit

## Kosten:

- **CAS:** je CHF 5'900 (ca. EUR 5'200)
- **DAS:** 3 CAS, gesamt CHF 17'700 (ca. EUR 15'700)
- **MAS:** 4 CAS + Masterarbeit (CHF 3'200), gesamt CHF 26'800 (ca. EUR 23'800)

# MAS Data Science



# CAS Information Engineering



- **Modul A: Scripting:**  
*Einführung Python; sciPy; scikit-learn; Datenextraktion, Datenanalyse und Datenvisualisierung*
- **Modul B: Datenbanken und Data Warehousing**  
*SQL; OLTP und OLAP; Data-Warehouse-Architektur; ETL-Prozesse*
- **Modul C: Information Retrieval**  
*Grundlagen Information Retrieval; Textanalyse; Sentiment-Analyse; mehrsprachiges Information-Retrieval; Multimedia Information Access*
- **Modul D: Big Data**  
*NoSQL; MapReduce mit Hadoop; Apache Spark*

# CAS Datenanalyse



- **Modul A: Grundlagen der Datenanalyse:**  
*Einführung in R und R-Studio; Explorative Datenanalyse und Grafiken; Einführung Wahrscheinlichkeitsrechnung und schliessende Statistik*
- **Modul B: Multiple Regression und Prognose**  
*Multiple Lineare Regression; Modellwahl und Diagnostik; Zeitreihen (STL-Zerlegungen, exponentielles Glätten, AR-Prozesse)*
- **Modul C: Clustering und Klassifikation**  
*Hauptkomponentenanalyse; Clustering; Klassifikationsverfahren*

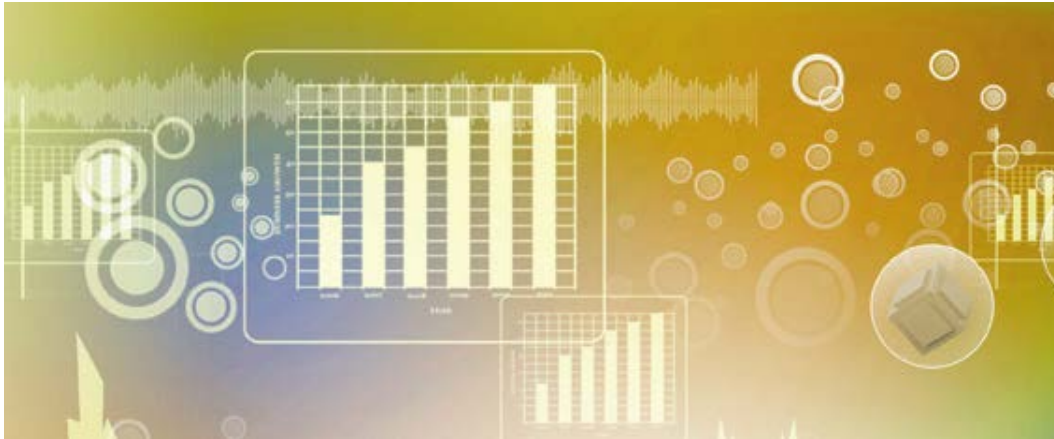
# CAS Machine Intelligence



- **Modul A: Machine Learning**  
*Grundlagen Machine Learning (Python); Clustering; Klassifikation; Anomaly Detection; Feature Engineering*
- **Modul B: Deep Learning**  
*Neuronale Netze; Neuronale Netze mit vielen Schichten; Autoencoder; Frameworks (TensorFlow); Architekturen (CNNs, RNNs); Aktuelle Entwicklungen*
- **Modul C: Text Analytics**  
*Grundlagen; Text-Klassifikation; Sentiment-Analyse, Keyword-Extraktion etc.*
- **Modul D: Advanced Topics in Big Data**  
*Vetiefung Apache Spark; Big-Data-Technologien für strukturierte und unstrukturierte Daten; Streaming-Daten; skalierbare ML-Algorithmen*



# CAS Statistical Modelling



- **Modul A: Information Processing with R**  
*Rmarkdown; Data Wrangling; Visualisierung; interaktive Browser-Interfaces (Shiny)*
- **Modul B: Advanced Regression Modelling**  
*Logistische Regression; GLMs; GAMs; Robuste und moderne Schätzmethoden*
- **Modul C: Analysis of Time to Event Data**  
*Modelle für Warte- und Ausfallzeiten; zensierte Daten; schliessende Statistik für Time-to-Event-Daten*
- **Modul D: Network Analysis**  
*Netzwerke und Graphen; deskriptive Analyse; stochastische Modelle; Diffusion und Epidemien*

# CAS Data Product Design



- **Modul A: Data-Specific Service Design**  
*Service Design; Service Science und Service Dominant Logic; Rapid Service Prototyping; Präsentation von Firmen-Cases*
- **Modul B: Data-Specific Business Model Design**  
*Business Model Design; Service Ecosystem Design; Quantifizierung*
- **Modul C: Praxis-Workshop**  
*Vertiefung und Anwendung der Konzepte*
- **Modul D: Datenschutz- und Datensicherheit**  
*Grundlagen Datenschutz und Datensicherheit; Rechtliche Aspekte und Ethik; Diskussion von Fallstudien*



# Weiterbildung

Grosse Nachfrage, **CAS Datenanalyse** und **CAS Information Engineering** werden inzwischen 2 mal pro Jahr angeboten (trotzdem lange vorher ausgebucht).

Unterricht durch interne Dozierende aus verschiedenen Fachgebieten:

- aus der **Statistik** (CAS Datenanalyse, CAS Statistical Modelling)
- aus der **Informatik** (CAS Information Engineering, CAS Machine Intelligence)
- **Operations Management** und **Jura** (CAS Data Product Design)

Dozierende unterrichten jeweils  $\frac{1}{2}$  - 1 Modul (Spezialisten!).

**Pluralismus bei den Programmiersprachen:** R und Python, weitere Software wie Hadoop, Spark etc.

**Leistungsnachweise** durch Prüfungen und / oder Projektarbeiten.



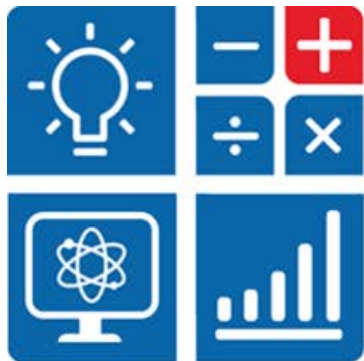
# Community

## Tagungen und Workshops

- Organisation der **Swiss Conference on Data Science (SDS)** 2014-2016 (seit 2017 organisiert von der Swiss Alliance for Data-Intensive Services)
- Konferenzen **Swiss Text** (jährlich seit 2016)
- **Deep Learning Day** seit 2016
- Zürich Machine Learning Meetup (T. Stadelmann, Co-Organisation 2015-2017)

## Weitere Aktivitäten

- **Datalab-Seminar:** “Brown-Bag-Seminar” zum informellen Austausch über Software, Methoden, Projekte und andere Data-Science-Themen (meist interne Referenten, manchmal Gäste)
- Gründungsmitglied im Nationalen Thematischen Netzwerk **Swiss Alliance for Data-Intensive Services** (mehrere Hochschulen und Unternehmen)



# data1ab

[www.zhaw.ch/data1ab](http://www.zhaw.ch/data1ab)