



FH Salzburg  
Informationstechnik &  
System-Management

# 2<sup>nd</sup> International Data Science Conference 2019

Conference Guide



Technology  
Health  
Media

## Inhalt

<b>Preface</b>	4 – 5
<b>Committees</b>	6
<b>The iDSC Idea</b>	7
<b>Program Overview</b>	8 – 11
<b>Social Event</b>	12
<b>Keynote Speakers</b>	14 – 15
<b>Workshops</b>	16 – 17
<b>Abstracts – Research Track</b>	18 – 22
<b>Conference Dinner</b>	23
<b>Venue &amp; Travel Information</b>	24 – 25
<b>Notes, Imprint</b>	26 – 27



# Welcome to Salzburg University of Applied Sciences!



**Ladies and gentlemen,  
dear colleagues,  
welcome to the 2<sup>nd</sup> International Data Science Conference.**

With the general topic "Data Science – Analytics and Applications", this year's conference refers to the fact that companies have already moved to apply data science, adding sustainable value to the potential benefits of digitalization in an agile environment.

The iDSC conference brings together researchers, scientists, engineers and entrepreneurs. We discuss new approaches in the fields of machine learning, artificial intelligence, data mining and visualization to establish efficient solutions to react flexibly and quickly to market conditions and customer requests in order to take into account changes that can be entrepreneurially decisive.

That's why data science has established itself as the cornerstone of institutions, organizations and companies and why it is crucial for analytical processes as well as for the realization of essential applications in business and research.

Take the chance to talk to scientists and business experts in the field of data science to discuss new ways within their respective domains. Participate in our workshops about state-of-the-art tools and topics and get inspired by our keynotes and talks from science and industry.

Our gratitude goes to our keynotes and industry speakers as well as our researchers, session chairs, review teams, to the program committee, to all participants and colleagues who made this event possible.

We wish you exciting days and many stimulating discussions.

Your conference chairs,



**Peter Haber**  
Salzburg University of Applied Sciences



**Thomas J. Lampoltshammer**  
Danube University Krems



**Manfred Mayr**  
Salzburg University of Applied Sciences

## Committees

### Conference Chairs

Peter Haber	Salzburg University of Applied Sciences
Thomas J. Lampoltshammer	Danube University Krems
Manfred Mayr	Salzburg University of Applied Sciences

### Local Organizers

Nicole Siebenhandl	Salzburg University of Applied Sciences
Maximilian Tschuchnig	Salzburg University of Applied Sciences
Dominik Vereno	Salzburg University of Applied Sciences

### Program Committee

David Anastasiu	San José State University
Arne Bathke	University of Salzburg
Markus Breunig	University of Applied Science Rosenheim
Frank Danielsen	University of Agder
Eric Davis	Industrial Labs
Günther Eibl	Salzburg University of Applied Sciences
Süleyman Eken	Kocaeli University
Karl Entacher	Salzburg University of Applied Sciences
Mohammad Ghoniem	Luxembourg Institute of Science and Technology
Elmar Kiesling	Vienna University of Technology
Michael Gadermayr	Salzburg University of Applied Sciences

Manolis Koubarakis	National and Kapodistrian University of Athens
Maria Leitner	Austrian Institute of Technology
Elena Lloret Pastor	University of Alicante
Giuseppe Manco	University of Calabria
Robert Merz	Vorarlberg University of Applied Sciences
Edison Pignaton de Freitas	Federal University of Rio Grande do Sul
Florina Piroi	Vienna University of Technology
Kathrin Plankensteiner	Vorarlberg University of Applied Sciences
Siegfried Reich	Salzburg Research
Peter Reiter	Vorarlberg University of Applied Sciences
Michael Ruzicka	Cockpit
Marta Sabou	Vienna University of Technology
Johannes Scholz	Graz University of Technology, Institute of Geodesy
Axel Straschil	pmlT Consult
Lórinç Thurnay	Danube University Krems
Andreas Unterweger	Salzburg University of Applied Sciences
Gabriela Viale Pereira	Danube University Krems
Stefan Wegenkittl	Salzburg University of Applied Sciences
Karl-Heinz Weidmann	Vorarlberg University of Applied Sciences
Anneke Zuiderwijk - van Eijk	Delft University of Technology

# The iDSC Idea

The conference gives the participants the opportunity, over the course of three days, to delve into the most current research and up-to-date practice in data science and data-driven business. The conference is split into two parallel tracks, a research track and an industry track.

## Research & Industry Track

The research track offers a series of short presentations from data science researchers regarding their current work in the fields of data mining, machine learning, data management and the entire spectrum of data science.

In the industry track, practitioners demonstrate showcases of data-driven business and how they use data science to achieve organizational goals, with a focus on market & trends, energy, manufacturing, quality assurance and health and sports.

## Workshops & Keynotes

Our sponsors will have their own, special platform via **workshops** to provide hands-on interaction with tools or to learn approaches towards concrete solutions. In addition, there will be an **exhibition** of the sponsors' products and services throughout the conference, with the opportunity for our participants to seek contact and advice.

Rounding out the program, we are proud to present **keynote presentations** from leaders in data science and data-driven business, both researchers and practitioners. These keynotes will be distributed over both conference days, providing times for all participants to come together and share views on challenges and trends in data science.

# Program Overview

WEDNESDAY  
22<sup>nd</sup> May 2019

	Research Track Room 012	Industry Track Room 013
08:00 am	Registration (Foyer)	
09:00 am	Opening and Welcome Room 013	
09:15 am	<b>Keynote</b> Josef Walzl (Amazon Web Services): From Data to Value in Industrial Use Cases Room 013	
10:00 am	<b>Keynote</b> Bodo Hoppe (IBM Germany): What Lies Ahead of Us? IBM's View on Technology Trends Room 013	
10:45 am	Break	
11:00 am	Room 012 <b>Data Analytics   Complexity</b> Chair: Lőrinc Thurnay  <b>11:00 Jad Rayes</b> (George Mason University): Exploring Insider Trading Within Hypernetworks  <b>11:30 Abdel Aziz Taha</b> (Research Studios Austria): Chance Influence in Datasets With Large Number of Features	Room 013 <b>Data Analytics   Market &amp; Trends</b> Host: Michael Ruzicka / Axel Straschil  <b>11:00 Guido Harucksteiner</b> (Skidata): Innovation & Trends with Special Regard to a FRUGAL Approach and Emerging Markets  <b>11:30 Robert Stubenrauch</b> (Business Upper Austria): From Industry 4.0 to Industrial Data - Accompanying Corporate Networks Through the Digital Age
12:00 am	Lunch	
01:15 pm	<b>Keynote</b> Peter Parycek (Danube University Krems): Data-Driven Policy-Making Room 013	
02:00 pm	<b>Keynote</b> Christian Blakely (PricewaterhouseCoopers Zurich): Real-Time Learning and Prediction in (Un)structured Data Room 013	
02:45 pm	Coffee Break	

	Research Track Room 012	Industry Track Room 013
03:30 pm	<b>Data Analytics   NLP and Semantics</b> Chair: Andreas Unterweger  <b>03:30 Hakkı Yağız Erdinç</b> (Donanim Haber): Combining Lexical and Semantic Similarity Methods for News Article Matching  <b>04:00 Martin Schnöll</b> (Fact AI): The Effectiveness of the Max Entropy Classifier for Feature Selection  <b>04:30 Lőrinc Thurnay</b> (Danube University Krems): Impact of Anonymization on Sentiment Analysis of Twitter Postings	<b>Data Analytics   Energy</b> Host: Michael Ruzicka / Axel Straschil  <b>03:30 David Steidl, Daniel Wagner</b> (Verbund): EDA & Anomaly Detection  <b>04:15 Norbert Walchhofer</b> (Cognify), <b>Stefanie            Kritznern</b> (Salzburg AG): No Space for Gaps - Forecasting Challenges in the Control Energy Market
05:00 pm	Closing	
07:00 pm – 10:30 pm	<b>Conference Dinner: Mozart Dinner Concert</b> <b>(only if booked in advance)</b>	

# Program Overview

THURSDAY  
23<sup>rd</sup> May 2019

	Research Track Room 012	Industry Track Room 013	Workshops Room 018
08:30 am	Registration (Foyer)		
09:30 am	<b>Keynote</b> Stefan Wegenkittl (Salzburg University of Applied Sciences): Confronting the Small Data Challenge: Achieving Successful Digital Transformation in SMEs Requires Transforming Processes, Roles and Technologies Room 013		
10:15 am	Break		
10:30 am	<b>Data Analytics   Modelling</b> Chair: Werner Pomwenger  <b>10:30 David Anastasiu</b> (San José State University): A Data-Driven Approach for Detecting Autism Spectrum Disorders  <b>11:00 Ioannis Gkioulekas</b> (University College London): Optimal Regression Tree Models Through Mixed Integer Programming  <b>11:30 Lina Stanzel</b> (AEE INTEC): A Spatial Data Analysis Approach for Public Policy Simulation in Thermal Energy Transition Scenarios	<b>Data Analytics   Manufacturing</b> Host: Michael Ruzicka / Axel Straschil  <b>10:30 Dubravko Dolic</b> (Continental): Predictive Maintenance for Tires  <b>11:00 Neha Sehgal</b> (University of Huddersfield & Valuechain Ltd.): 'Champions' & 'Strugglers' of UK Manufacturing Sector - The Tale from Open Data  <b>11:30 Sven Ahlinder</b> (Volvo): Visualization of many variables, by projection onto two of them	<b>10:30 Steadforce:</b> Candy Consumption Forecasting
12:00 am	Lunch		
01:15 pm	<b>Keynote</b> David Anastasiu (San José State University): The AI Data Revolution: Doing More With Less Data Labeling Room 013		
02:00 pm	Break		

	Research Track Room 012	Industry Track Room 013	Workshops Room 018
02:15 pm	<b>Data Analytics   Comprehensibility</b> Chair: Johannes Scholz  <b>02:15 Maciej Skorski</b> (Dell): Probabilistic Approach to Web Waterfall Charts  <b>02:45 Shefali Virkar</b> (Danube University Krems): Facilitating Public Access to Legal Information - A Conceptual Model for Developing an Agile Data-driven Decision Support System  <b>03:15 Wolfgang Kremser</b> (Salzburg Research): Do We Have a Data Culture?	<b>Data Analytics   Quality Assurance</b> Host: Michael Ruzicka / Axel Straschil  <b>02:15 Oskar Preinfalk, Rene Leikermoser</b> (Spar ICS): HADES – Anomaly detection in retail processes  <b>03:00 Bernhard Redl, Simon Stiebellehner</b> (craftworks): CI/CD for Machine Learning	<b>02:15 Cognify:</b> Profit from Prophet
03:45 pm	Coffee Break		
04:30 pm	<b>Short Papers</b> Chair: Thomas J. Lampoltshammer  <b>04:30 Dejan Radovanovic</b> (Salzburg University of Applied Sciences): Neural Machine Translation from Natural Language into SQL with state-of-the-art Deep Learning Methods  <b>05:00 Sebastian Malin</b> (Fachhochschule Vorarlberg): Smart Recommendation System to Simplify Projecting for a HMI/SCADA Platform  <b>05:30 Maximilian E. Tschuchnig</b> (Salzburg University of Applied Sciences): Adversarial Networks – A Technology for Image Augmentation  <b>06:00 Melanie Zumtobel</b> (Fachhochschule Vorarlberg): Using Supervised Learning to Predict the Reliability of a Welding Process	<b>Data Analytics   Health &amp; Sports</b> Host: Michael Ruzicka / Axel Straschil  <b>04:30 Jonathan Boidol, Stephan Schiffner</b> (Steadforce): Discovering and Extracting Knowledge from large Text Collections - Introduction to Text mining on the example of clinical trials  <b>05:15 Richard Mohr</b> (Tehedge): Big Data and Real Time Analytics in US Football Sports. Implementing the perfect match day with Arena Analytics for the San Francisco 49ers.  <b>06:00 Konrad Linner, Philipp Lukas</b> (Solvistas): Ice Hockey and Data Science - How the EHC Black Wings Linz wants to become a champion with the help of Data Science	<b>04:30 MathWorks:</b> Predictive Maintenance with MATLAB
06:30 pm	Networking Room 017		
08:00 pm	Closing		

## Social Event

Friday, 24<sup>th</sup> May 2019

# Half-Day Panorama Tour

To round off the conference, we will do a guided bus tour through Salzburg's lake district »Salzkammergut« on Friday, May 24<sup>th</sup> 2019. You can continue the discussion in Austria's picturesque scenery.

Meeting time: 8:45 am  
Paris-Lodron-Straße, Salzburg  
(opposite »Hotel am  
Mirabellplatz«)  
Departure: 9 am  
Return to Salzburg: 1 pm

We spend half a day in the famous towns Fuschl and St. Gilgen (30-45 min. stay) alongside some of the most beautiful lakes in Austria. You have the opportunity to reflect on presented topics and converse with attendees and speakers alike.

The tour is included in the conference ticket (only if separately booked in advance).



# 2<sup>nd</sup> International Data Science Conference 2019

## Program Details

## Keynote Speakers

**David C. Anastasiu**  
San José State  
University



David C. Anastasiu is an assistant professor in the Department of Computer Engineering at San José State University. His research interests fall broadly at the intersection of machine learning, data mining, computational genomics, and high performance computing. He was awarded the Next Generation Data Scientist (NGDS) Award at the 2016 IEEE International Conference on Data Science and Advanced Analytics.

**Christian D. Blakely**  
Pricewaterhouse-  
Coopers Zurich



Christian started his career at NASA Goddard Space Flight Center in Washington DC as an atmospheric physicist. After completing a Ph.D. in Computational Science at University of Maryland, he had a 3 year long deep dive into big data and machine learning while doing a post-doctoral fellowship with the United States Department of Commerce. He moved to Switzerland to pursue a career in FinTech, and now leads the machine learning team for PwC Switzerland, focusing on real-time machine learning technologies. He is also a part time classical concert pianist.

**Bodo Hoppe**  
IBM Germany



Bodo Hoppe is a distinguished engineer in IBM Research & Development GmbH located in Germany. He is responsible for the overall design quality and functionality of the IBM z Systems microprocessor and system ASICs. He is a thought leader in verification methodology and has introduced multiple innovations to verification approaches. He is leading the agile transformation enabling the user experience driving the hardware development from design thinking to feature based development. He is involved in multiple activities and collaborations with IBM Research on new technologies.

## Keynote Speakers

**Peter Parycek**  
Danube University  
Krems



Peter Parycek is full professor for e-governance, head of the Department for E-Governance and Administration at the Danube University Krems and head of the Competence Centre Public IT at Fraunhofer Fokus Berlin, funded by the Ministry of Interior. In August 2018 he has become a member of the digital council to advise the German government on the important matter of digitalisation.

**Josef Waltl**  
Amazon Web  
Services



Josef Waltl leads the global partner ecosystem for industrial software at Amazon Web Services (AWS). Prior to AWS he worked in Siemens building on software strategy and mergers & acquisitions for product lifecycle management, smart grid and mobility. He holds a Ph.D. and an MBA from the Technical University Munich as well as a Dipl.-Ing. in Computer Science from the University of Salzburg and a Dipl.-Ing. from the Salzburg University of Applied Sciences in Telecommunications Engineering and Systems.

**Stefan Wegenkittl**  
Salzburg University  
of Applied Sciences



Stefan Wegenkittl is academic programme director of the Applied Image and Signal Processing degree programme as well as senior lecturer and head of the department of applied mathematics and data-mining at Salzburg University of Applied Sciences. There, he also heads the Applied Data Science Lab which conducts research in the areas medical image processing, biosignal processing and natural language processing. Current research is on various aspects of machine learning, representation learning and feature extraction in the aforementioned areas of application.

# Workshops

## Workshop 1

Thursday, 23<sup>rd</sup> May 2019 – 10:30 am

Steadforce: »Candy Consumption Forecasting«

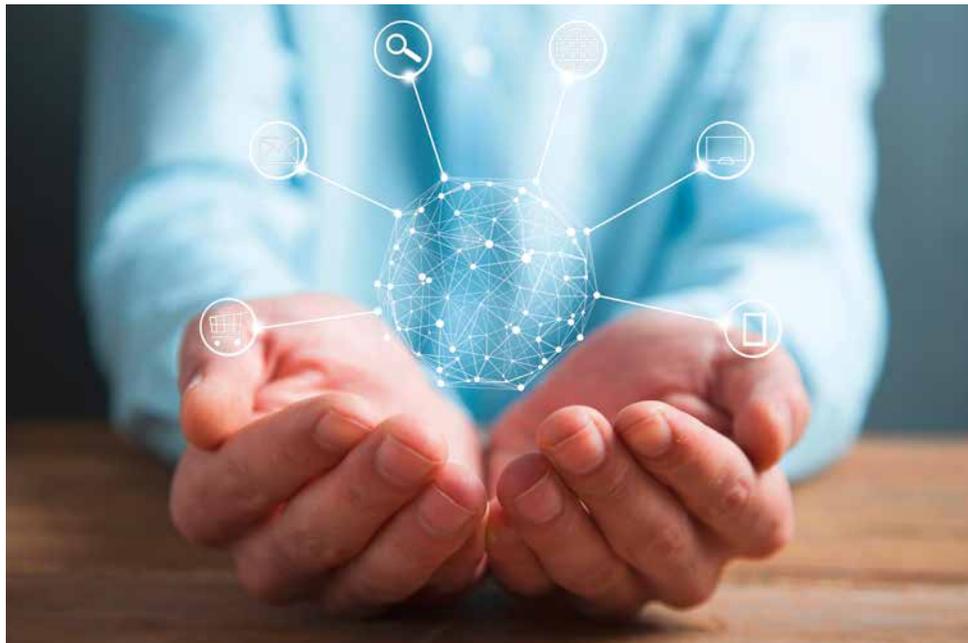
This workshop is targeted at students and professionals who are interested in advanced data science methods and best practices.

The aim of the workshop is to develop a predictive model for the demand for sweets in an office environment. Using this in-house example, data cleaning, preprocessing and modelling will be carried

out. Classical outlier detection and filtering methods are applied, and as prediction models we will use survival regression.

The participants have the opportunity to follow a complete data science use case on the basis of an unusual data set.

Development environment of the workshop: Python, Jupyter and the Python Ecosystem.



## Workshop 2

Thursday, 23<sup>rd</sup> May 2019 – 02:15 pm

Cognify: »Profit from Prophet«

Providing high quality forecasts timeseries like sales prices can be quite complex considering influential factors like holidays, seasonality effects etc.

Prophet is an open source forecasting tool developed by Facebook and an intuitive approach to modeling such timeseries data.

During the workshop we will give an introduction to timeseries forecasting with Prophet and other timeseries approaches in theory and hands-on practice.

## Workshop 3

Thursday, 23<sup>rd</sup> May 2019 – 04:30 pm

MathWorks: »Predictive Maintenance with MATLAB «

In this workshop, an application from the field of predictive maintenance is used to demonstrate how data analytics turn data to decisions. In contrast to preventive maintenance, which follows a set timeline, predictive maintenance schedules are determined by analytic algorithms and data from equipment sensors. With predictive maintenance, organizations can identify issues before equipment fails, pinpoint the root cause of the failure, and

schedule maintenance as soon as it is needed. The workshop covers the complete workflow from accessing and preprocessing data through developing predictive models using machine learning techniques and visualizing results to deploying the final algorithms in production systems and embedded devices.

# Abstracts – Research Track

## Data Analytics | Complexity

### Jad Rayes and Priya Mani (George Mason University): Exploring Insider Trading Within Hypernetworks

Insider trading can have crippling effects on the economy and its prevention is critical to the security and stability of global markets. It is hypothesized that insiders who trade at similar times share information. We analyze 400 companies and 2,000 insiders, identifying interesting trading patterns in these networks that are suggestive of illegal activity. Insiders are classified as either routine or opportunistic traders, allowing us to concentrate on well timed and highly profitable trades of the latter. Using trade classification and analyzing each trader's role in a hypernetwork, reveals cliques of opportunistic and routine traders. This idea forms the basis of a graph based detection algorithm that seeks to identify traders belonging to opportunistic cliques. The ideas of trade classification and trading cliques present interesting opportunities to develop more robust policing systems which can automatically flag illegal activity in markets, and predict the likelihood that such activity will occur in the future.

### Abdel Aziz Taha, Alexandros Bampoulidis and Mihai Lupu (Research Studios Austria): Chance influence in datasets with large number of features

Machine learning research, e.g. genomics research, is often based on sparse datasets that have very large numbers of features, but small samples sizes. Such configuration promotes the influence of chance on the learning process as well as on the evaluation. Prior research underlined the problem of generalization of models obtained based on such data. In this paper, we deeply investigate the influence of chance on classification and regression. We empirically show how considerable the influence of chance such datasets is. This brings the conclusions drawn based on them into question. We relate the observations of chance correlation to the problem of method generalization. Finally, we provide a discussion of chance correlation and guidelines that mitigate the influence of chance.

## Data Analytics | NLP and Semantics

### Mehmet Umut Sen, Hakki Yagiz Erdinc, Burak Yavuzalp and Murat Can Ganiz (Donanim-Haber, Sabanci University; DonanimHaber, Dogus University; DonanimHaber, Istanbul Technical University; VeriUs, Marmara University): Combining Lexical and Semantic Similarity Methods for News Article Matching

Matching news articles from multiple different sources with different narratives is a crucial step towards advanced processing of online news flow. Although, there are studies about finding duplicate or near-duplicate documents in several domains, none focus on grouping news texts based on their events or sources. A particular event can be narrated from very different perspectives with different words, concepts, and sentiment due to the different political views of publishers. We develop novel news document matching method which combines several different lexical matching scores with similarity scores based on semantic representations of documents and words. Our experimental result show that this method is highly successful in news matching. We also develop a supervised approach by labeling pairs of news documents as same or not, then extracting structural and temporal features. The classification model learned using these features, especially temporal ones and train a classification model. Our results show that supervised model can achieve higher performance and thus better suited for solving above mentioned difficulties of news matching to opportunistic cliques. The ideas of trade classification and trading cliques present interesting opportunities to develop more robust policing systems which can automatically flag illegal activity in markets, and predict the likelihood that such activity will occur in the future.

### Martin Schnöll, Cornelia Ferner and Stefan Wegenkittl (Fact AI GmbH; Salzburg University of Applied Sciences): The Effectiveness of the Max Entropy Classifier for Feature Selection

Feature selection is the task of systematically reducing the number of input features for a classification task. In natural language processing, basic feature selection is often achieved by removing common stop words. In order to more drastically reduce the number of input features, actual feature selection methods such as Mutual Information or Chi-Squared are used on a count-based input representation. We suggest a task-oriented approach to select features based on the weights as learned by a Max Entropy classifier trained on the classification task. The remaining features can then be used by other classifiers to do the actual classification. Experiments on different natural language processing tasks confirm that the weight-based method is comparable to count-based methods. The number of input features can be reduced considerably while maintaining the classification performance.

### Thomas J. Lampoltshammer, Lőrinc Thurnay and Gregor Eibl (Danube University Krems): Impact of Anonymization on Sentiment Analysis of Twitter Postings

The process of policy-modelling, and the overall field of policy-making are complex and put decision-makers in front of great challenges. One of them is present in form of including citizens into the decision-making process. This can be done via various forms of E-Participation, with active/passive citizen-sourcing as one way to tap into current discussions about topics and issues of relevance towards the general public. An increased understanding of feelings behind certain topics and the resulting behavior of citizens can provide great insight for public administrations. Yet at the same time, it is more important than ever to respect the privacy of the citizens, act in a legally compliant way, and therefore foster public trust. While the introduction of anonymization in order to guarantee privacy preservation represents a proper solution towards the challenges stated before, it is still unclear, if and to what extent the anonymization of data will impact current data analytics technologies. Thus, this research paper investigates the impact of anonymization on sentiment analysis of social media, in the context of smart governance. Three anonymization algorithms are tested on Twitter data and the results are analyzed regarding changes within the resulting sentiment. The results reveal that the proposed anonymization approaches indeed have a measurable impact on the sentiment analysis, up to a point, where results become potentially problematic for further use within the policy-modelling domain.

## Data Analytics | Modelling

### **Manika Kapoor and David Anastasiu (San Jose State University): A Data-Driven Approach for Detecting Autism Spectrum Disorders**

Autism spectrum disorders (ASDs) are a group of conditions characterized by impairments in reciprocal social interaction and by the presence of restricted and repetitive behaviors. Current ASD detection mechanisms are either subjective (survey-based) or focus only on responses to a single stimulus. In this work, we develop machine learning methods for predicting ASD based on electrocardiogram (ECG) and skin conductance (SC) data collected during a sensory challenge protocol (SCP) in which the reactions to eight stimuli were observed from 25 children with ASD and 25 typically developing children between 5 and 12 years of age. The length of the time series makes it difficult to utilize traditional machine learning algorithms to analyze these types of data. Instead, we developed feature processing techniques which allow efficient analysis of the series without loss of effectiveness. The results of our analysis of the protocol time series confirmed our hypothesis that autistic children are greatly affected by certain sensory stimulation. Moreover, our ensemble ASD prediction model achieved 93.33% accuracy, which is 13.33% higher than the best of 8 different baseline models we tested.

### **Ioannis Gkioulekas and Lazaros Papageorgiou (University College London): Optimal Regression Tree Models through Mixed Integer Programming**

Regression analysis is a tool for predicting the output variables from a set of known independent variables. Through regression, a function that captures the relationship between the variables is fitted to the data. Tree regression models are popular in the literature due to their ability to be computed quickly and their simple interpretations. However, creating complex tree structures can lead to overfitting the training data resulting in a poor predictive model. This work introduces a tree regression algorithm that employs mathematical programming to optimally split data into two sub regions, called nodes, and a statistical test to assess the quality of partitioning. A number of publicly available literature examples have been used to test the performance of the method against others that are available in the literature.

### **Lina Stanzel, Johannes Scholz and Franz Mauthner (AEE - Institut für Nachhaltige Technologien; Graz University of Technology, Institute of Geodesy): A Spatial Data Analysis Approach for Public Policy Simulation in Thermal Energy Transition Scenarios**

The paper elaborates on an approach to simulate the effect of public policies regarding thermal energy transition pathways in urban communities. The paper discusses the underlying methodologies of calculating Heating Energy demand of buildings and the rationale for potential zones for thermal energy systems. In order to simulate the effects of public policies on communities the authors developed a spatial Agentbased Model, where the buildings are the main objects that are subject to change, based on a number of both technically and socio-demographic parameters. In order to fill a spatial Agentbased Model with data a number of open source and commercially available datasets need to be spatially analyzed and merged. The initial results of the spatial Agent-based Model simulation show that public policies for thermal energy transition can be simulated accordingly.

## Data Analytics | Comprehensibility

### **Maciej Skorski (DELL): Probabilistic Approach to Web Waterfall Charts**

The purpose of this paper is to propose an efficient and rigorous modeling approach for probabilistic waterfall charts illustrating timings of web resources, with particular focus on fitting them on big data. An implementation on real-world data is discussed, and illustrated on examples. The technique is based on non-parametric density estimation, and we discuss some subtle aspects of it, such as noisy inputs or singular data. We also investigate optimization techniques for numerical integration that arises as a part of modeling.

### **Shefali Virkar, Chibuzor Udokwu, Anna-Sophie Novak and Sofia Tsekeridou (Danube University Krems; INTRASOFT International S.A.): Facilitating Public Access to Legal Information: A Conceptual Model for Developing an Agile Data-driven Decision Support System**

The European legal system is multi-layered and complex, and large quantities of legal documentation have been produced since its inception. This has significant ramifications for European society, whose various constituent actors require regular access to accurate and timely legal information, and often struggle with basic comprehension of legalese. The project focused on within this paper proposes to develop a suite of usercentric services that will ensure the real-time provision and visualisation of legal information to citizens, businesses and administrations based on a platform supported by the proper environment for semantically annotated Big Open Legal Data (BOLD). The objective of this research paper is to critically explore how current user activity interacts with the components of the proposed project platform through the development of a conceptual model. Model Driven Design (MDD) is employed to describe the proposed project architecture, complemented by the use of the Agent Oriented Modelling (AOM) technique based on UML (Unified Modelling Language) user activity diagrams to develop both the proposed platform's user requirements and show the dependencies that exist between the different components that make up the proposed system.

### **Wolfgang Kremser and Richard Brunauer (Salzburg Research Forschungsgesellschaft): Do we have a Data Culture?**

Nowadays, adopting a »data culture« or operating »data-driven« are desired goals for a number of managers. However, what does it mean when an organization claims to have data culture? A clear definition is not available. This paper aims to sharpen the understanding of data culture in organizations by discussing recent usages of the term. It shows that data culture is a kind of organizational culture. A special form of data culture is a data-driven culture. We conclude that a data-driven culture is defined by following a specific set of values, behaviors and norms that enable effective data analytics. Besides these values, behaviors and norms, this paper presents the job roles necessary for a datadriven culture. We include the crucial role of the data steward that elevates a data culture to a data-driven culture by administering data governance. Finally, we propose a definition of data-driven culture that focuses on the commitment to data-based decision making and an ever-improving data analytics process. This paper helps teams and organizations of any size that strive towards advancing their – not necessarily big – data analytics capabilities by drawing their attention to the often neglected, non-technical requirements: data governance and a suitable organizational culture.

Program  
23<sup>rd</sup> May 2019

## Short Papers

### Maximilian Ernst Tschuchnig (Salzburg University of Applied Sciences): Adversarial Networks — A Technology for Image Augmentation

A key application of data augmentation is to boost state-of-the-art machine learning for missing values and to generate more data from a given dataset. Additional to transformations or patch extraction as augmentation methods, adversarial networks can be used in order to learn the probability density function of the original data. Generative adversarial networks (GANs) are an adversarial method to generate new data from noise by pitting a generator against a discriminator and training in a zero-sum game trying to find a Nash Equilibrium. This generator can then be used in order to convert noise into augmentations of the original data. This short paper shows the usage of GANs in order to generate fake faces as well as tips to overcome the notoriously hard training of GANs.

### Dejan Radovanovic (Salzburg University of Applied Sciences): Neural Machine Translation from Natural Language into SQL with state-of-the-art Deep Learning methods

Reading text, identifying key ideas, summarizing, making connections and other tasks that require comprehension and context are easy tasks for humans but training a computer to perform these tasks is a challenge. Recent advances in deep learning make it possible to interpret text effectively and achieve high performance results across natural language tasks. Interacting with relational databases through natural language enables users of any background to query and analyze a huge amount of data in a user-friendly way. This paper summarizes major challenges and different approaches in the context of Natural Language Interfaces to Databases (NLIDB). A state-of-the-art language translation model developed by Google named Transformer is used to translate natural language queries into structured queries to simplify the interaction between users and relational database systems.

### Melanie Zumtobel, Kathrin Plankensteiner (FH Vorarlberg): Using supervised learning to predict the reliability of a welding process

In this paper, supervised learning is used to predict the reliability of manufacturing processes in industrial settings. As an example case, lifetime data has been collected from a special device made of sheet metal. It is known, that a welding procedure is the critical step during production. To test the quality of the welded area, End-of-Life tests have been performed on each of the devices. For the statistical analysis, not only the acquired lifetime, but also data specifying the device before and after the welding process as well as measured curves from the welding step itself, e.g., current over time, are available. Typically, the Weibull and log-normal distributions are used to model lifetime. Also in our case, both are considered as an appropriate candidate distribution. Although both distributions might fit the data well, the log-normal distribution is selected because the ks-test and the Bayesian Factor indicate slightly better results. To model the lifetime depending on the welding parameters, a multivariable linear regression model is used. To find the significant covariates, a mix of forward selection and backward elimination is utilized. The t-test is used to determine each covariate's importance while the adjusted coefficient of determination is used as a global Goodness-of-Fit criterion. After the model that provides the best fit has been determined, predictive power is evaluated with a non-exhaustive cross-validation and sum of squared errors. The results show that the lifetime can be predicted based on the welding settings. For lifetime prediction, the model yields accurate results when interpolation is used. However, an extrapolation beyond the range of available data shows the limits of a purely data-driven model.

### Sebastian Malin, Kathrin Plankensteiner, Robert Merz, Reinhard Mayr, Sebastian Schöndorfer, Mike Thomas (FH Vorarlberg, COPA-DATA GmbH): Smart recommendation system to simplify projecting for an HMI/SCADA platform

Modelling and connecting machines and hardware devices of manufacturing plants in HMI/SCADA software platforms is considered time-consuming and requires expertise. A smart recommendation system could help to support and simplify the tasks of the projecting process. In this paper, supervised learning methods are proposed to address this problem. Data characteristics, modelling challenges, and two potential modelling approaches, one-hot encoding and probabilistic topic modelling, are discussed.

Side Event – only if booked in advance  
Wednesday, 22<sup>nd</sup> May 2019

# Conference Dinner

07:00 – 10:30 pm,  
St. Peter Stiftskulinarium Salzburg

Enjoy a unique culinary and musical experience in the baroque hall of Europe's oldest restaurant: A delicious three-course-meal combined with some of Mozart's most enchanting compositions.

St. Peter Stiftskulinarium  
Sankt-Peter-Bezirk 1/4  
5020 Salzburg

#### Program:

Doors open, drinks served

First course

Arias & duets from »Don Giovanni«

Main course

Arias & duets from »Le Nozze di Figaro«

Dessert

»A Little Night Music«

Arias & duets from »The Magic Flute«



# Venue & Travel Information

**Conference Venue**  
 FH Salzburg  
 Campus Urstein  
 Urstein Süd 1  
 5412 Puch/Salzburg



## Arrival & Location plan

A location plan as well as information on how to get to the conference venue (public transport, car) and parking sites can be found at [www.idsc.at/location/arrival-information](http://www.idsc.at/location/arrival-information).

Please note that there is only a limited amount of parking sites available at Urstein Campus. When arriving by car, please only park in the appropriate sections.

If possible use public transport (the train »S3« directly stops at the campus! Get off at station »Puch Urstein«). The regional bus lines 160 and 165 also stop directly at the campus.

## Taxi & public transport Salzburg

Taxi: +43 662 8111  
 Public Transport Salzburg:  
[www.salzburg-verkehr.at](http://www.salzburg-verkehr.at)

## Online Registration

The online registration for participating in the 2<sup>nd</sup> International Data Science Conference 2019 at Salzburg University of Applied Sciences is possible until May 15, 2019.

[www.idsc.at/registration](http://www.idsc.at/registration)

## Documentation of the iDSC 2019

The responsible organizer of the event will take photos and make videos in the course of the conference (incl. audio tracks) in its prevailing interests of documentation and publication of the event and its contents. Through participating in the event, the data protection policy ([www.idsc.at/privacy-policy](http://www.idsc.at/privacy-policy)) is acknowledged and Salzburg University of Applied Sciences is authorized to use the above mentioned photos and video footage (incl. audio tracks) without monetary compensation and without any kind of local, temporal and content related restrictions.

## Tips for hotels in Salzburg

### Motel One – Salzburg Mirabell

Elisabeth Kai 58-60, 5020 Salzburg  
 T +43 662 885200  
[www.motel-one.com](http://www.motel-one.com)  
[salzburg-mirabell@motel-one.com](mailto:salzburg-mirabell@motel-one.com)

### Mercure Hotel – Salzburg City

Bayerhamerstrasse 14 a, 5020 Salzburg  
 T +43 662 8814380  
[www.mercure.com](http://www.mercure.com)  
[H0984@accor.com](mailto:H0984@accor.com)

### Hotel Lasserhof

Lasserstrasse 47, 5020 Salzburg  
 T +43 662 873388  
[www.lasserhof.com](http://www.lasserhof.com)  
[info@lasserhof.com](mailto:info@lasserhof.com)

### Wyndham Grand

Fanny-von-Lehnert-Straße 7, 5020 Salzburg  
 T +43 662 46880  
[www.wyndhamgrandsalzburg.com](http://www.wyndhamgrandsalzburg.com)  
[info@wyndhamgrandsalzburg.com](mailto:info@wyndhamgrandsalzburg.com)





## Sponsors



### **Cognify KG**

Intelligent algorithms, complex  
data analysis and technology  
consulting

[www.cognify.ai](http://www.cognify.ai)



### **The MathWorks GmbH**

Mathematical computing software

[www.mathworks.de](http://www.mathworks.de)

## Supported by



**ITG: innovative consulting and  
location development**

Salzburg's innovation centre

[www.itg-salzburg.at](http://www.itg-salzburg.at)