Data Science is not yet automated. Let’s get back to work!

As you may have read in the last issue of Swiss Analytics Magazine, Data Science is not (yet) automated. We still need Data Scientists to develop, test and implement Predictive Analytics within companies. Our magazine is there to support you in this journey to data-driven decision making.

In this release, Marcel Baumgartner explains how Analytics can be used for (internal) audit. Jérôme Berthier describes what is “The Voices” category within Data Scientists. Jacques Zuber provides an introduction to Deep Learning. Karyn Murphy lists 8 common mistakes in IoT. Additionally, I reviewed the book - let’s call it a Bible - Business Forecasting.

I take the opportunity to remind you of our main platforms of communication:

- Meetup (www.meetup.com/swiss-analytics): Analytics events proposed by our association.

- LinkedIn (www.linkedin.com/groups/4586163): Our discussion group (news, job offers, events, etc.)

- Website (www.swiss-analytics.com): Content we share (magazine, event slides, etc.)

As discussed during our latest General Assembly (Feb. 1st 2017), both the existing Committee and President have been re-elected. We also decided to stop paid membership for the association. Thanks to our sponsors, major running costs are covered and the committee can now focus on strategic tasks such as the organization of events and magazines. To become a free member, simply join our LinkedIn group.

Sandro Saietta, Data Scientist
Demand Planning
at Nespresso

www.swiss-analytics.com
DATA ANALYTICS FOR INTERNAL AUDIT

Introduction

Large publicly listed companies not only have external auditors who check the books, but often also a large community of internal auditors. These collaborators provide the company with a sufficient level of assurance in terms of adherence to internal and external rules and guidelines. This covers financial aspects (spend, invoices, investments...), human resources (working time, payroll...) but also production related aspects (e.g. food safety and quality).

One of the strongest trends observed in internal auditing communities is the more and more widespread use of Data Analytics. The term refers to the use of data, statistical methods and statistical thinking as a way of working, in addition to traditional auditing methods like interviews, document and process reviews, etc. This is naturally a trend not only observed in audit: many other business processes now rely more and more on data-driven decision making, and this manifests itself in the buzz around “Big Data”, “Data Science” and “Business Analytics”.

In this article, we describe different approaches to ensure that Data Analytics is used efficiently in a large company for controlling and internal audit.

The Promise of Data Analytics

The main promise of data analytics is coverage. While 10 or 15 years ago, it was necessary to create a sample of financial documents in order to find potential issues, this is typically not needed anymore. Modern business software solutions (e.g. like those from SAP) allow the extraction of all financial documents, and therefore provide the basis for an exhaustive analyses of all of them. The efficiency gain is huge.

Line of Defense

Internal audit is considered to be the 3rd line of defense in most companies. The 2nd line is typically provided by the functions (e.g. Sales, Marketing, Finance, Supply Chain, Manufacturing...), who build compliance by design into processes and software solutions. In addition, companies may have an Internal Control department, which provides top down control mechanisms and analyses. The 1st line of defense is the operational management: through a cascading structure, they implement the control procedures and supervise the employees. Data Analytics is particularly relevant for both the 2nd and 3rd line of defense.

Top-down vs. Bottom-up

The analytics that are carried out to identify potential issues in business processes can be run top-down or bottom-up. Top-down refers to a global or regional organization, which runs scripts on all categories and geographies on a regular basis (monthly, quarterly), for selected, high risk processes. They then share the outcome with the local process owners, and ensure that proper actions are taken. Internal control organizations typically work like this. The advantage is clearly that all the risky processes are covered globally, and therefore the level of assurance is rather high. However, the nature of these checks make them also somehow “simple”: the algorithms behind such top-down typically don’t use advanced statistical techniques, and are essentially rule-based. The objective is to provide a clearly understandable output, whilst reducing the false positive rate as much as possible.

1 See this paper for more details: http://tinyurl.com/nStgazt
Bottom-up up data analytics refers to scripts and algorithms that are run by internal auditors, ad-hoc, within the scope of their audit mission. With such a framework, a company can develop more sophisticated scripts, using modern statistical methods like clustering and classification, or using graph networks, in order to find issues that nobody has seen before. But clearly, it will be difficult to do this on like a global or regional level.

At Nestlé, we do both: more top-down for the 2nd line of defense, and more bottom-up for the 3rd line of defense.

**From “Data to Analytics” to “Analytics to Data”**

The community of statisticians, data miners, or data scientists, always worked under the principle “Data to Analytics”. To e.g. run a clustering algorithm in R or SAS, the data first need to be obtained from the source system in the form of a text file, and then be imported into the statistical software. This has naturally limitations, as soon as the data sources are too large. And it takes time to download, prepare and upload the data, and this is rarely done in one single iteration.

Since a few years, new technology provides the possibility to run the analytics directly on the source database, with immense performance improvements: in these “in-memory” systems (like SAP HANA), an algorithm running for 24 hours previously can take a few minutes. At Nestlé, this journey has started: we have proven the feasibility e.g. to move in memory data from the source SAP HANA system into a R server, and then send the results back. Very little time is spent on the data transfer, and there is no need any more for downloads and interfaces: we have a seamless integration.

However, this generates other issues. Now suddenly, complex analytical algorithms run directly on the live system, potentially impacting business operations. Additionally, it is not allowed to develop algorithms directly on the live database: the code first needs to be written and tested on development systems, and then transported carefully into the production environment. However, the test systems don’t contain real data. These algorithms do generate false positives, and it is not possible to estimate the rate of them using test data. The truth on the performance of the algorithms is only revealed on real data, but then it’s too late to adapt the algorithm, and the development cycle starts again.

Therefore, the IS/IT organizations in such companies will need to develop other processes to ensure that data scientists can develop their code efficiently, running short cycles of development, testing, corrections, on systems using real data and having a similar computing performance.

**Bottom-Up Is Driving Innovation**

The internal audit organization at Nestlé has built its data analytics strategy strongly around the bottom-up approach. This is also referred to as “Self-Service Analytics”. That is the internal auditors are empowered via training, coaching, support and software solutions to run most of their analytics on their own.

The internal audit organization at Nestlé has a Data Analytics team, who has the mission to provide the framework so that internal auditors can take full advantage of the data the company has. The team works in close collaboration with the auditors to not only coach and train them, provide clear documentation, but also to innovate and develop new scripts regularly.

In the recent years, we have been able to generate valuable insight into financial and food production processes through the use of statistical and graphical methods. Here are two examples:
• The financial documents behind the Accounts Payables (we buy materials and services from suppliers, are invoiced, and we pay) and Accounts Receivables (we obtain orders from customers, we invoice them, and we get paid) processes need deep controls and investigations. During an entire year, a mid-sized business can generate hundreds of thousands such documents. Internal control has developed numerous rule-based methods to identify documents that need investigation. In order to find red flags that are not identified by the rules, we have started to use bottom up hierarchical clustering. The difficulty was to develop the dissimilarity matrix between the financial documents, as these documents are characterized by both categorical (type of document, who created it, by which process…) and numerical variables (the amount, debit or credit). Gower’s metric, as available in the function “daisy” of the R library “cluster”, solved this for us. Now internal auditors have a sample of say 50 documents out of 100’000, that are markedly different from all the “normal” ones, and they can start building the story of the document, and eventually decide whether it is a real red flag or a false positive. If something went wrong, chances are high that the issue will become visible in the sample.

• We know how much of a specific semi-finished product H1 was consumed in the finished product F1, and we also know how much H1 consumed of the raw material R1. We have this information for all pairs of products in a given factory, over a period of say one year. The question now is: can we develop a more optimal algorithm to determine which finished materials are using a specific raw material? Representing the data as a graph network helped us greatly. The nodes are the materials, and the edges are given by the consumption matrix described above. Within this graph, we can then find the shortest paths between raw and finished materials, and therefore obtain exactly the information we are looking for. In the illustration below, R1 is used in F4, as there is a path from R1 through H1 and H2 to F4:

The R package “igraph” provides all the necessary functions to build the graphs, and to run the analytics (shortest path, distance, degrees of nodes…).
Outlook

Data-driven controlling and auditing will further accelerate, there’s no doubt about this. However, there are challenges. One is the needed mindset change of IS/IT organizations, to ensure that analytics can be developed and tested in a much faster way compared with traditional change management processes. The other challenge is the human resources: the statisticians / data scientists that develop the routines. These people not only need a deep knowledge of statistical methods, they also need to understand the business needs, i.e. having the ability to translate a controlling idea into a piece of code that works and is relevant. Outsourcing the coding might not be an option, therefore mastering programming languages like R, SAS, SQL... to build the algorithms are also needed. The competition for these people will be fierce in the coming years. Those companies that provide them with interesting problems, large access to data and an environment where they can collaborate across functions and geographies will succeed.

About the author

Marcel Baumgartner works for Nestlé since 1994, in its headquarters in Vevey, Switzerland. Nestlé is the world’s leading Nutrition, Health and Wellness company. He has a diploma as an applied mathematician from the “Ecole Polytechnique Fédérale de Lausanne” (www.epfl.ch), Switzerland, and a masters in Statistics from Purdue University in West Lafayette, IN, US. He was the global lead for Demand Planning Performance and Statistical Forecasting, and since November 2014, he focuses on providing Data Analytics capabilities for internal auditors. He is also the president of the Swiss Statistical Society (www.stat.ch).
The commonly called deep learning or hierarchical learning is now a popular trend in machine learning. Recently during the Swiss Analytics Meeting Prof. Dr. Sven F. Crone presented how we can use deep learning in the industry in a forecasting perspective (beer forecasting for manufacturing, lettuce forecasting in retail outlets, container forecasts). Deep learning has a variety of applications as for example image and handwritten character recognition. It analyses a picture and will be able to conclude if it is a dog, a human or something else. After a learning process, deep learning first understands your handwriting and then can read and interpret a draft paper you have quickly written. But briefly what is exactly deep learning?

In the artificial intelligence process, deep learning plays an important role. It is considered as a method of machine learning and roughly speaking means neural networks. More precisely artificial neural networks are intended to simulate the behaviour of biological systems composed of multiple layers of nodes (or computational units), usually interconnected in feed-forward way. Each node in one layer has directed connections to the nodes of the subsequent layer. Feed-forward neural networks can be considered as a type of nonlinear predictive models that takes inputs (very often huge amount of both labelled and unlabelled data), transforms and weights these through plenty of hidden layers to produce a set of outputs (predictions). The use of a sequence of layers, organised in deep or hierarchical levels, explains the term of « deep learning ». Each layer receives as input the information contained in the previous layer, transforms it to the following layer and of course complete and improve it.

We consider the well-known numerical image recognition problem to illustrate how deep learning works in practice. In the first hidden layer, the network analyses the pixels and classifies them, for example by colour. Obtained results are then studied in the second layer for identifying relevant relationships. For instance, some lines and shadowing effects are detected. A third hidden layer analyses and combines these curves for discovering forms such as human faces. New layers can be
added to improve and refine the deep learning model for discovering better patterns. This process can continue until the network generates as output a desirable image where the nature of the picture can be identified (a dog, a cat or a human person for example).

To conclude this academic tip, deep learning is a learning mechanism. It is very attractive and effective for almost any task of machine learning and Internet of Things (IoT) especially for classification. But it needs in fact a lot of data and requires very long times for model training, especially when the number of hidden layers is large. Nevertheless, the availability of new hardware, particularly GPUs, and modern parallel computing have made computations much cheaper and faster.

Neural network models are very flexible but typically over-parametrized. They are so-called « black-box » models and provide results which are not always human-interpretable even for an expert.

Recent developments have been carried out to improve deep learning methods and algorithms. Different libraries are now available as for example the open-source TensorFlow developed by Google and MXNetR, darch, deepnet, h2o and deепr, libraries of the free statistical computing system R. Deep Learning is in commercial business software packages too as for example in SAS Enterprise Miner.
INTERVIEW OF JÉRÔME BERTHIER, HEAD OF BI AND BIG DATA AT ELCA

1. Can you tell us who you are and how you came to the field of Data Science?

My name is Jerome Berthier, I am an engineer in Computer Science and I have an MBA in management. After 10 years working in different roles for an IT provider (developer, sales representative, managing director), I joined ELCA in 2012 to head the BI division. At that time Big Data was starting to become a mainstream concept in IT and I had the great opportunity to be in charge of developing ELCA’s expertise in this field.

We had to start from scratch, as Big Data was completely new to us, and so, to research the subject we engaged an expert in NLP and created our own Big Data Lab devoted to testing and evaluating algorithms and Big Data solutions available on the market, basically anything that could help us to better understand the principles of Big Data and how they can be applied. This research was an amazing experience which achieved excellent results.

Since then, I have kept up to date on the evolution of Big Data and have been active in raising awareness of its benefits, through presentations in various contexts: for Elca customers, at IT events and in media interviews.

2. On your LinkedIn profile, you describe yourself as “A Voice” of Data Science. What do you mean by that?

To be honest, I don’t believe that any individual data scientist can be skilled in all fields: IT, marketing, communication, maths, sales, statistics, business (banking, insurance, travel...) and so on: an All-in-One data scientist, so to speak. Of course, if you know one, I would be delighted to meet them and have them in my team.

The applications of Data Science are often highly specialized and there are several types of data scientists. The key is to take advantage of these different profiles and involve them together to create a strong team of data scientists. SAS institute defines 7 basic profiles:

- **The Geeks 41%**: The Geeks are the largest group in our sample and have the largest female membership of all the groups at 37 per cent. They have a naturally technical bias, strong logic and analytical skills. Essentially “black and white” thinkers, they like to speak plainly and stick to the point – don’t expect them to be moved by emotionally charged arguments. With their attention to detail and fondness for the rules, the Geeks are well suited to roles such as defining systems requirements, designing processes and programming.

- **The Gurus 11%**: The next largest group, the Gurus, has a measure of reactive introversion, like the Geeks, which predisposes them to scientific and technical subjects. Yet they also display a diametrically opposite characteristic: the strong presence of proactive extroversion, including solid and often highly persuasive communications and social skills. The Gurus can play a very important role by using their enthusiasm, tact and diplomacy to promote the benefits of the data sciences to those holding the purse strings, or who have the authority to give projects the green light.

1 http://tinyurl.com/p2aleku
• The Drivers 11%: The Drivers are proactive introverts: highly pragmatic individuals who use their determination and focus to realise their goals. Self-confident and results-oriented, they are ideal project managers and team leaders, who excel at prioritising, monitoring and driving projects to a successful conclusion.

• The Crunchers 11%: This category is probably one of the least self-promoting groups. Strongly reactive – rather than proactive – personalities, the Crunchers like routine and constancy. They display high technical competence and consistency, making them superb in a range of technically-oriented support roles including data preparation and entry, statistical analysis, monitoring of incoming data and quality control.

• The Deliverers 7%: Like the Drivers, these individuals are proactive and well suited to project and man management. This is also the group with the largest proportion of men at 80 per cent. However, the Deliverers also have a strong pre-disposition towards acquiring and/or applying technical skills. So, while they are capable of bringing focus and momentum to ensure project success, they are also likely to understand the finer technical details and devise solutions in much greater technical depth.

• The Voices 6%: The Voices are strong communicators with less apparent detailed technical knowledge than the Gurus. The presence of this group suggests a strong demand for natural promoters who have the ability to generate enthusiasm for the potential of big data and the data sciences at a conceptual level – rather than the practical or technical level. The Voices are strongly valued for their positive outlook, and may be engaged in presenting the results of big data projects as well as supporting their implementation.

• Other Personalities 13%: A smaller number of respondents displayed a range of other traits.

• The Ground Breakers: offer new approaches, new methods and new possibilities, drawn from a mix of inspiration and dogged logical thinking. Roles include: system design and algorithm development.

• The Seekers: combine superb technical knowledge and understanding with inquisitiveness and a drive to find solutions. Roles include: research.

• The Teachers: skilled at imparting knowledge and inspiring others to want to learn. Roles include: training and mentoring.

• The Lynchpins: important team players who may not have a depth of technical knowledge but provide essential support services. Roles include: coordination and administration.

So I am a “Voice”. Why did I decide to use this term?

Often our customers prefer to start with POC or POV to evaluate the potential of Big Data. But even if the results show great promise, it is still not easy to get the necessary budget from the Board (C-level) to go a step further, because past IT solutions did not necessarily live up to expectations. So it is my role to accompany the teams in their POC in order be familiar with the company context and to assist them in presenting the results to C-levels in such a way as to encourage continued support...

3. What do you see as the main skills that a Data Scientist needs in 2017?

All of them, of course , but I see 3 of them as being especially important:

• Big Data has changed the relationship between infrastructure and analytics. In the past, there were 2 silos: on one hand, analytics; and on the other hand, infrastructure. Now it’s impossible to do useful analysis if you don’t understand the underlying architecture of
the infrastructure; and it’s impossible to size the infrastructure accurately if you don’t know what type of analysis will be used.

So knowledge of Big Data infrastructure will be as important as ever.

• Content analytics is a high priority for me because more and more projects use flow aggregation, Chabot, document analysis, email analysis …

• Finally, complex event processing is becoming increasingly relevant.

4. Based on your personal experience, what is the biggest challenge you face when involved in Data Science projects?

For me, there are 2 key challenges:

• First of all, Data is our “raw material”. Without it we are out of work. I often read and hear that the world is now submerged in data… That may be true but how much of this data is really relevant?

I have seen a number of projects in which important data is lacking (which seems incredible when we realize how much data we possess…) or/and to which important data is not accessible. Quality/governance of data is clearly a big problem everywhere.

• Another problem is habit. How can we explain to people who have spent years working in the same way that everything must change drastically because all of the old habits have now become obsolete?

I once did a customer segmentation project which showed that the existing segmentation was irrelevant because not based on data analysis.

5. If you could give just one piece of advice to a future data scientist, what would it be?

Open a book on Business Intelligence.

I often receive new data scientists who know nothing about BI, ETL, SQL language or even what a database is…

I consider Big Data to be an evolution of our classic BI, but the basic principles stay the same: collection of data, transformation, merging, analysis and finally decision-making based on the results.

Today unstructured data can be used, merging can be on the fly and decisions can be predictive or prescriptive, but the principles remain the same.

Furthermore, outside of Academic circles, each time you will want to work with big data, you will have to deal with old BI and old sources of data. So it’s necessary to understand how the old system works, too.
BOOK REVIEW: “BUSINESS FORECASTING: PRACTICAL PROBLEMS AND SOLUTIONS”

Reviewed by Sandro Saitta, Data Scientist, Demand Planning at Nespresso

The book Business Forecasting is a compilation of existing articles. Gilliland, Tashman and Sglavo edited what can be named the Bible of Business Forecasting. It contains around 50 articles that appeared in journals such as Foresight and Journal of Business Forecasting. The scope of the book is wide with topics ranging from forecast accuracy and forecastability, to new product forecasting and S&OP. Business Forecasting is structured in 4 categories:

- Fundamentals
- Methods
- Performance & Reporting
- Process & Politics

Articles are usually short and well written. They allow a good overview of the discussed topic and several references are provided to deep dive when needed. For a few topics, more than one article is available. The reader gets various points of view, making the book more comprehensive. Only in the performance evaluation section, the introduction of some articles are redundant. While a few articles are to be read as food for thought, most of them are directly usable in your daily forecasting activities.

Even though all the book is valuable to me, the section on process and politics of forecasting is definitely the one with the most added value for the practitioner. The book only contains published material. The question of its relevance is to be asked. For newcomers in the field of forecasting, as well as people without subscription to magazines such as Foresight, the book is a must-have. For the others, while it is still nice to have such a compilation in one place, the book added value is lower. Let’s be honest however: there is no such comprehensive and up-to-date gathering of business forecasting articles out there. Do not hesitate too much and just go for it, you won’t be disappointed.
IOT SUCCESS DEPENDS ON DATA GOVERNANCE, SECURITY AND PRIVACY
8 COMMON MISTAKES SPOTLIGHT PITFALLS AND LESSONS LEARNED

Karyn Murphy, Managing Editor, IDG Strategic Marketing Services

Tyrone Grandison, Deputy Chief Data Officer at the US Department of Commerce, says “agility and nimbleness” are key to capitalizing on emerging trends like the internet of things (IoT). But he warns that “waterfall demands” on the data management life cycle will force firms to increase their focus on more complex and detailed data governance. These intensified demands, in turn, lower the probability of organizational success in this space. When industry experts were asked to weigh in on the topic, several key themes emerged. The result is a list of 10 common mistakes organizations have made in their efforts to seize the momentum around IoT.

Engaging innovation without the proper preparation

We all know we shouldn’t innovate for innovation’s sake. According to Sam Edelstein, Chief Data Officer at the City of Syracuse, “The biggest danger relative to IoT likely has to do with treating it as a shiny object.” He warns that preparation will “take time and money,” which affects its potential. “Without it,” he adds, “investment in IoT could result in little or no benefit.”

Not clearly understanding data or accurately defining the problem

Preparation must be rooted in clarity. Dana Blouin, IoT thought leader, reports that the biggest mistake a company could make to put themselves in jeopardy would be to not clearly define a data privacy policy, which clearly outlines what the scope of the data being collected will be and how it will be used.

Not asking the right questions

Kirk Borne, Principal Data Scientist at Booz Allen Hamilton, reminds us that “the promise of IoT is greater visibility and actionability.” Further, he contends that “this promise can go unfulfilled” if we don’t ask the right questions, such as:

- Who owns the data?
- How do we ensure data quality, discovery, usability and security for the many different teams and business units that create, use and manage the data?
- What are the key business questions and goals that are driving what data we collect and use?
- How do we manage ad hoc data analytics? Do we restrict it or encourage it?

Failing to anticipate the flood of data and breadth of complexity

“Data governance is getting evermore complex with the devolution of company tech borders. [with] data flowing to and from a breadth of devices, which are more often mobile and across a wider range of operating systems and platforms,” says Ian Moyse, Sales Director at Axios Systems.
Keeping too strong of a hold on data

To many, governance is synonymous with control – that is, to ensure security and compliance, even data quality. But Chuck Martin, Editor at MediaPost, contends that it’s a mistake to “too tightly restrict the ebb and flow of information, especially among and between connected devices.” That kind of grip can quickly break down the agility and nimbleness for which we strive.

Underestimating security and privacy implications

Bill McCabe, Recruiter at SoftNet, says it’s a mistake to “underestimate the added security and privacy demands that come with the introduction of IoT.” When everything is connected, everything is at risk, Gilmore points out, and “a data breach destroys consumer trust and can devastate an enterprise’s reputation and business. In the rush to capitalize on big data and connected devices, businesses are putting themselves at big risk by not putting privacy and security first. “IoT data that’s collected for analytics often includes sensitive customer information. That data should be treated like an asset – governed, secured and safeguarded for privacy.

Taking data platforms (and people skills) for granted

We often assume that the infrastructure – both the platform and the people – are there to support our innovations. But Steele urges us to take a hard look at such assumptions. “The infrastructure that data lands on is overlooked and is a critical foundation for dealing with mass amounts of data in an efficient and secure way. Companies must utilize scale-out or hyperconverged infrastructure with native security and Hadoop Distributed File System (HDFS) capabilities.” Then, he says, we can put greater focus on quality and governance. Moyse points to “strong, in-house expertise and change methodologies ... It will be all too easy for an employee or department to self-serve an IoT device, as we have seen with bring-your-own-device (BYOD) and departmental cloud applications.” This, he says, exposes the business to “risk from an unplanned and unexpected IP device connecting through the corporate network.”

Operating under a veil of complacency

We need to take ownership of our data governance. Martin warns that we mustn’t fall prey to “complacency regarding external technologies, expecting that outside suppliers have products that are completely bulletproof.” Do the necessary due diligence to make sure governance is covered – or pay the price.
Governments and businesses are expecting the Internet of Things (IoT) to hit the mainstream by 2020. Early adopters have already gained some experience. 75 teams were interviewed on recent experiences across all industries.

4 KEY LESSONS

- **Scope is almost always underestimated**
- **User experience defines results**
- **Data & governance skills are critical**
- **Maturity dictates value**

DIVERSE EXPECTATIONS

- **43%** Improved operational efficiency
- **36%** Improve user experience
- **29%** New product or service design
- **25%** Improve resource management

3 MOST POPULAR LESSONS

- **Apply design thinking**
- **Test in advance**
- **Plan for scalability**

CLOSING THE SKILLS GAP

- Storytelling/data flow visualisation
- Interpreting results
- Analysing data
- Critical thinking/pragmatic data scientist

Read the full report at [www.sas.com/iotbook](http://www.sas.com/iotbook) to learn more.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2016, SAS Institute Inc. All rights reserved. 1463549UK0916
The 4th Swiss Conference on Data Science, SDS|2017, is going to be held on the 16th of June 2017 at Kursaal, Bern, Switzerland. The conference brings together opinion leaders, practitioners, decision makers and researchers with interest in Data Science and related fields of study.

SDS|2017 is the 4th in the SDS conference series which was started in 2014 by the research group Datalab of the Zurich University of Applied Sciences. Which started as a half day workshop with national speakers and participants in the year 2014, have now grown to a full day conference including multiple tracks with international speakers and participants. The highlights of previous years SDS conferences can be seen in the official page of the conference.

SDS|2017 is being organised by the newly founded Swiss Alliance for Data-Intensive Services (Data+Service alliance). The Data+Service alliance is a Swiss National Thematic Network that targets the creation of business value from data through new industrial services and new digital and mobile services. It connects companies with academic institutions from the areas of data science and service science with the goal of producing business innovation.

In the past, SDS has presented talks in the area of cutting edge data science technology, concepts and trends, with focus on big data, data warehousing, artificial intelligence, machine learning, deep learning and more. The goal of the conference is to foster the exchange of ideas specifically for the Swiss market and to sustain the community of data scientists in Switzerland.

SDS|2017 invites and encourages data scientists as well as enthusiasts, from industry and academia to participate and share their experiences with the community. We invite talks in the areas of data product design, business models, applications of artificial intelligence, machine learning, data warehousing, analytics, cyber security, laws and ethics in data science, demonstration of technology for big/small data and more.

SDS|2017 is proud to have some of the world renowned data scientists and analysts participate as keynote speakers. The talks focus on the social implications of data science, some of which are beneficial and positive whereas some misleading and undesirable. The talks reflect upon the social issues arising with data science, how data can lead to inferences that were not meant to be retrieved or how we should have access to the data we produce.

1 www.sds2017.ch
2 www.zhaw.ch/datalab
3 www.zhaw.ch/datalab/sds2016
4 http://www.data-service-alliance.ch/
We have the honour of having the following data scientists as the keynote speakers.

Claudia Perlich, Dstillery, USA

Clemens Cap, University of Rostock, Germany

David Kriesel, Procter & Gamble, Germany

Gregory Grefenstette, Florida Institute for Human & Machine Cognition, France

The conference registration is open until the 8th of June 2017, we invite you to participate and strengthen further the vibrant and growing community of data scientists/enthusiasts in Switzerland.

Write to us at info@sds2017.ch should you have any questions regarding the sponsorship opportunities, program, exhibition etc. We look forward to meeting you at SDS|2017.
SWISSTEXT 2017: THE NATIONAL CONFERENCE FOR TEXT ANALYTICS

Mark Cieliebak, Conference Chair of SwissText 2017 (other roles: Lecturer at Zurich University of Applied Sciences (ZHAW) and CEO of SpinningBytes AG)

SwissText is a one-day conference on automatic text analytics, which takes place June 9, 2017 in Winterthur. The national conference gives an overview of existing solutions and technologies in automatic text understanding/natural language processing/computational linguistics.

The conference brings together experts from all areas of text analytics: practitioners and researchers, data scientists and software vendors, machine learning experts and end users. In several keynotes and technical talks, it will cover the most interesting topics in the field: how to build a Chatbot, what is state-of-the-art in machine translation, or which types of texts can be generated automatically.

First Edition with Overwhelming Success

The first edition of the conference was held in 2016, with more than 170 participants it was much larger than expected – in fact, a change of venue was necessary to host all these experts. The distinguished lineup of speakers included experts from international IT giants such as Google and IBM, Swiss companies (e.g. SwissCom), local startups, and several Swiss universities. Overall, there were over 30 keynotes, presentations and posters.

Broad Support by Swiss Community

This year’s conference is again strongly supported by the analytics community: it is organized by the Datalab of Zurich University of Applied Sciences (ZHAW), more than 10 Swiss universities and research associations are official partners of the conference, and it is sponsored by SpinningBytes, a private company for data and text analytics.

Peter Stengard, Principal Data Scientist at Microsoft, already agreed to give a keynote at SwissText 2017, and more speakers will be announced soon.

Presentations for the Audience

The conference will see the following types of presentations: Surveys, which describe current trends and solutions from “a bird eyes view”; Technical Papers for new and innovative results in all aspects of text analysis; Showcases, which describe interesting and successful projects and applications; and Demonstrations of tools and products for automatic text analytics.

Get Involved

Presentations can be submitted until April 10, 2017.

Participation fee for SwissText 2017 is CHF 130,- and includes admission to all sessions, lunch and after-conference Apéro. Registration is open until May 15, 2017.

For more information on SwissText 2017, visit the conference website at www.swisstext.org.
SUPPORT THE SWISS ASSOCIATION FOR ANALYTICS!

If you like our association and would like to contribute, you can write an article for our magazine, be a speaker at one of our events or start a discussion on our LinkedIn group:

LinkedIn: www.linkedin.com/groups/4586163
E-mail: info@swiss-analytics.com
Meetup: www.meetup.com/swiss-analytics

As a non-profit organization, we would also welcome any donation. Your participation will help us cover regular costs (web hosting, meetup subscription, etc.)

Our bank account is:

Name: Swiss Association for Analytics
Address: Chemin des Fontannins 12, 1066 Epalinges
Bank: PostFinance
Account: 85-122140-9
IBAN: CH65 0900 0000 8512 2140 9
BIC: POFICHBEXXX

Thanks for your support!

---

IMPRESSUM

The Swiss Analytics Magazine is proposed by the Swiss Association for Analytics (www.swiss-analytics.com). The magazine is printed by Gasser Media (www.gasser-media.ch). Articles are written by experts in analytics. The views expressed in this magazine do not necessarily represent the views of the Swiss Association for Analytics. It is not allowed to reproduce, duplicate, copy, sell, resell or exploit any portion of the magazine without agreement of the Swiss Association for Analytics (info@swiss-analytics.com).

Printed in April 2017, Le Locle, Switzerland
LE MEILLEUR DE VOS DONNÉES !
COLLECTER, ORGANISER, DIFFUSER.

GASSER MEDIA

ÉDITION
GESTION DE PROJET
DEPUIS 2001

IMPRESSION
PRODUCTION À LA DEMANDE
DEPUIS 1947

PUBLICATION
WEB, APP, CROSS-MEDIA
DEPUIS 2004

ADRESSE
Gasser Media SA
Jambe-Ducommun 6a
CH-2400 Le Locle

CONTACT
Tél. +41 32 933 00 33
info@gasser-media.ch
www.gasser-media.ch

Pour suivre nos actualités :
www.gasser-media.ch/blog