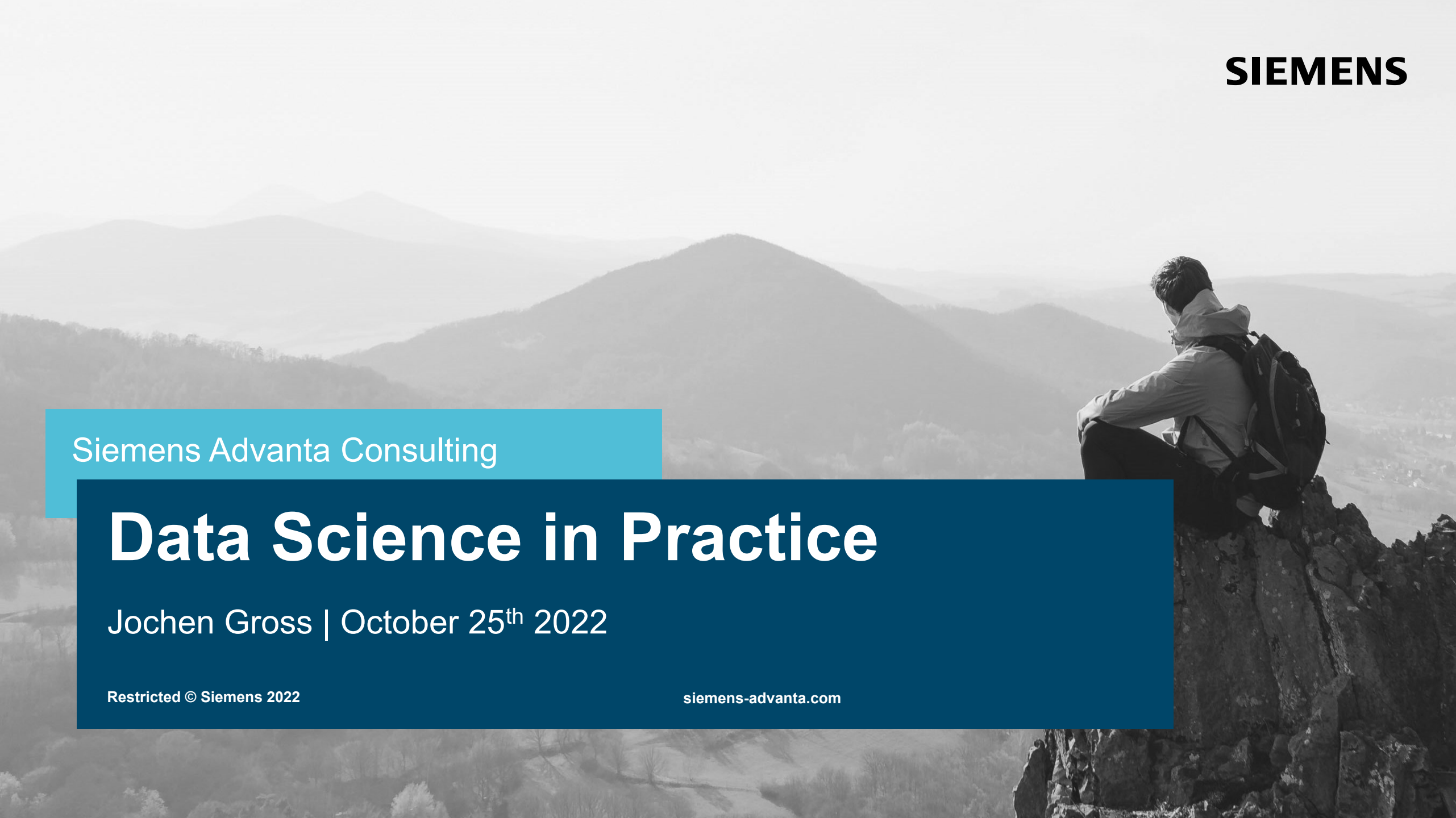
SIEMENS



Siemens Advanta Consulting

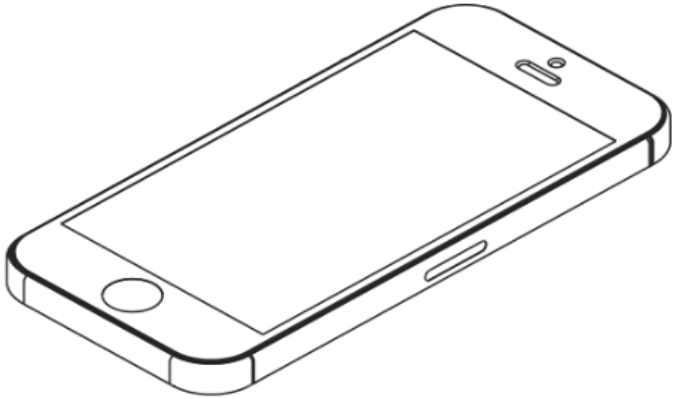
Data Science in Practice

Jochen Gross | October 25th 2022

siemens-advanta.com

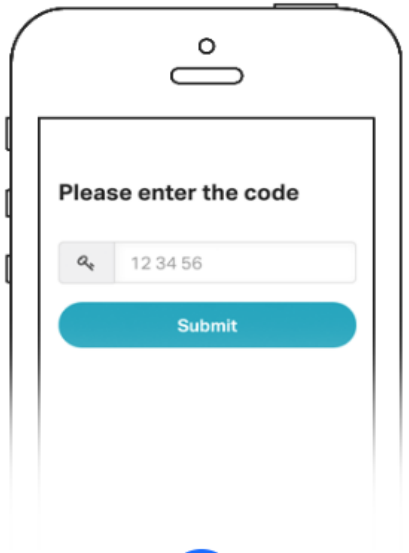
Go to www.menti.com and use the code 4957 5929

Let's get to know each other & discuss the topics you are interested in today



1

2

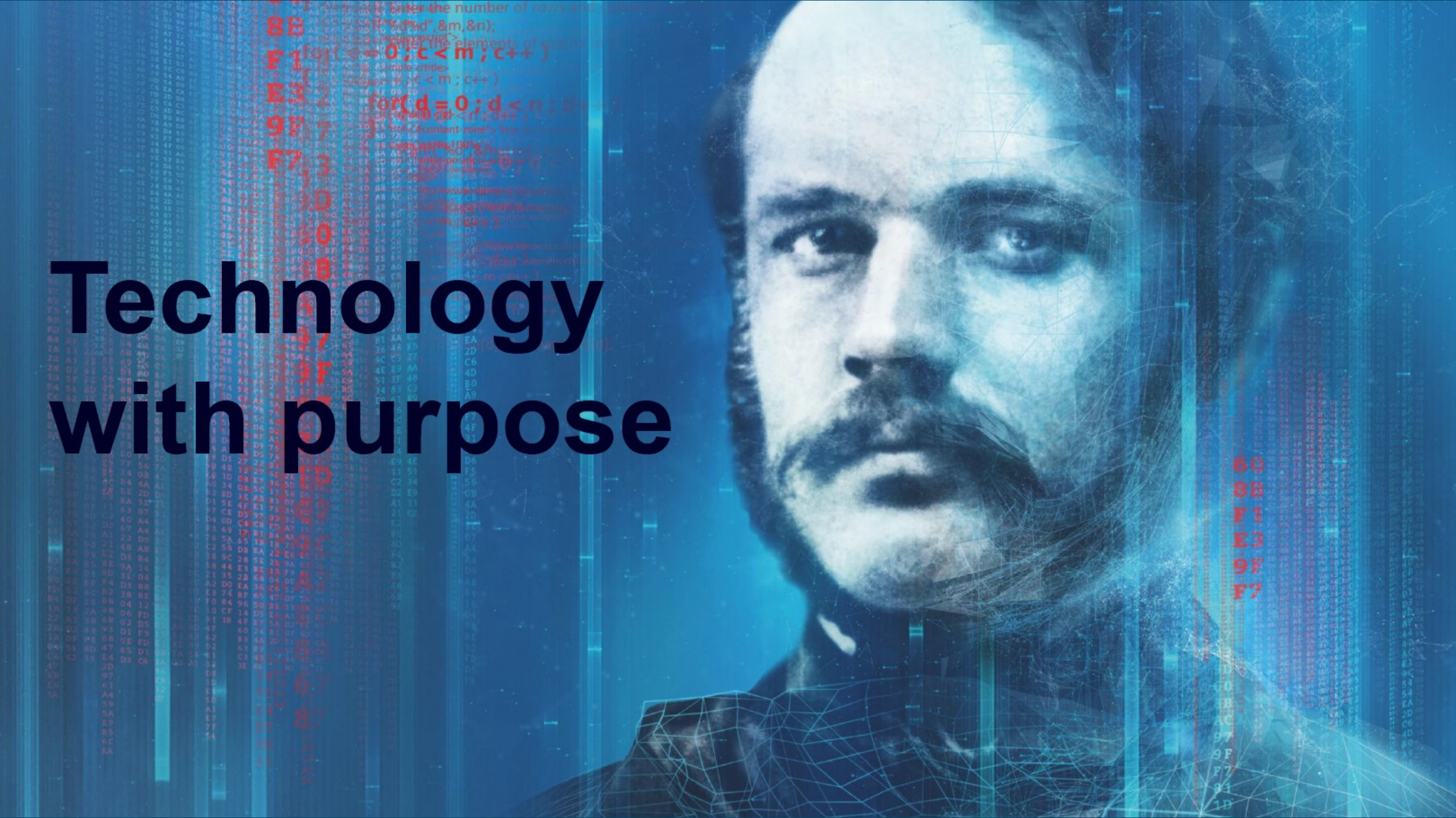


3

In addition, please use the chat function to raise your hand or ask questions directly during the talk!

SIEMENS

Technology with purpose



Siemens innovations

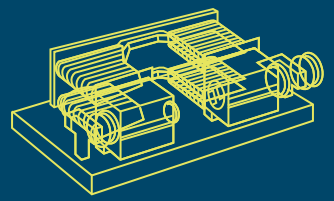
Milestones in a 173-year history

1816 – 1892

Company founder, visionary and inventor

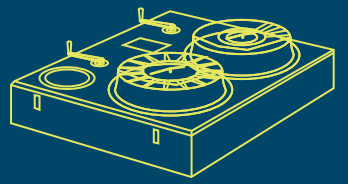
1866

The dynamo makes electricity part of everyday life



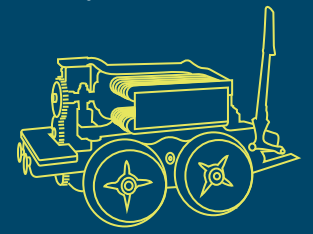
1847

Pointer telegraph lays the foundation of Siemens as a global company



1879

World's first electric railway

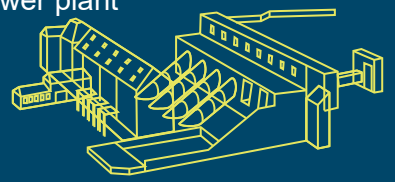


Source: Siemens Advanta Consulting

SIEMENS

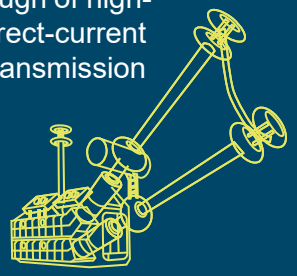
1925

Siemens electrifies the Irish Free State with a hydroelectric power plant



1975

Breakthrough of high-voltage direct-current (HVDC) transmission



2010

TIA Portal takes automation a stage further

2016

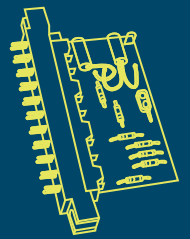
MindSphere, the cloud-based IoT operating system

2019

Launch of first project for Siemensstadt 2.0

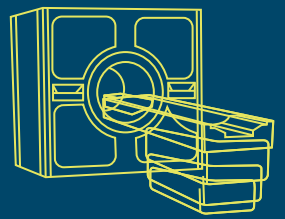
1959

SIMATIC revolutionizes automation



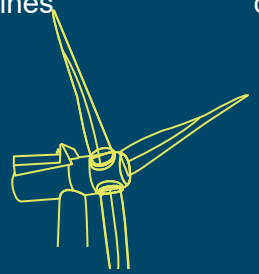
1983

First magnetic resonance imaging scanner



2012

Test operation of the world‘s largest rotor for offshore wind turbines



2018

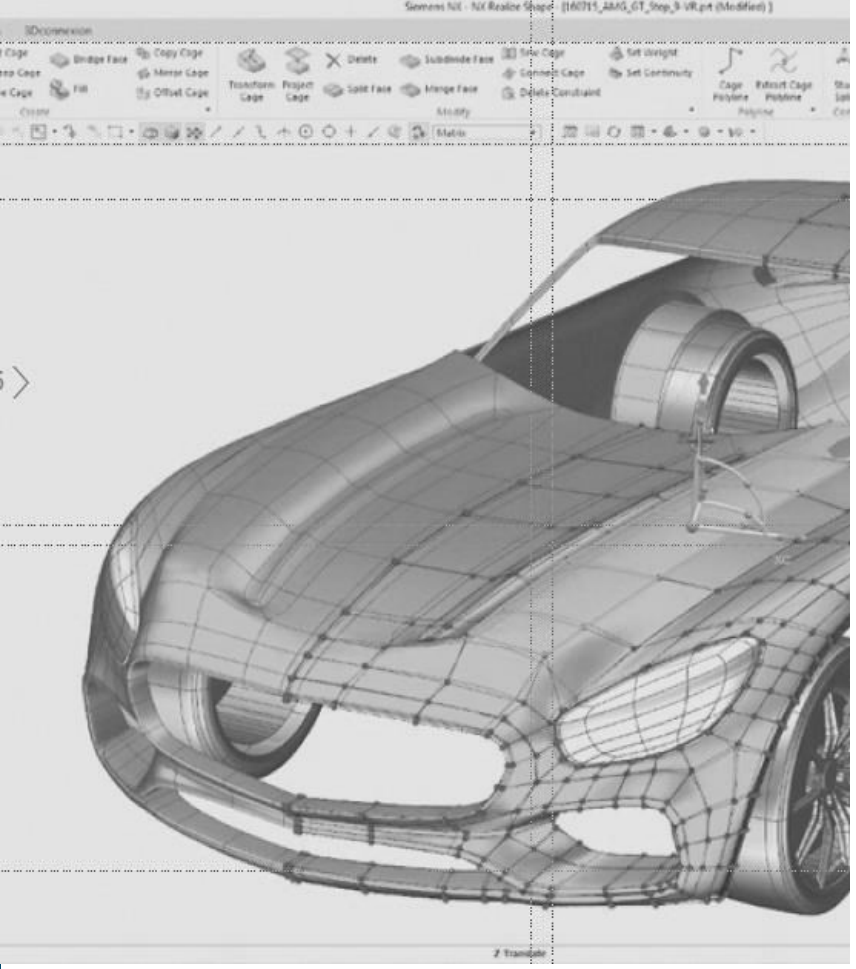
Charter of Trust: a joint initiative for a secure digital world

2020

Comfy workplace app makes it safe to return to the office during the coronavirus pandemic

Source: Siemens Advanta Consulting

SIEMENS



We commute in cars, designed with **Siemens Software …**

… built in **factories automated by Siemens …**

… and charged with renewable and decentral **Siemens Smart Grid.**

Source: Siemens Advanta Consulting

SIEMENS

Setup of Siemens AG

Businesses

Digital Industries

Smart Infrastructure

Mobility

Siemens Advanta

Portfolio Companies

Siemens Healthineers[1]

Countries

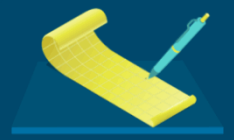
Service & Governance

1 Publicly listed subsidiary of Siemens; Siemens' share in Siemens Healthineers: 75%

SIEMENS

Siemens Advanta is a global consulting and professional service group with a strong focus on digitalization

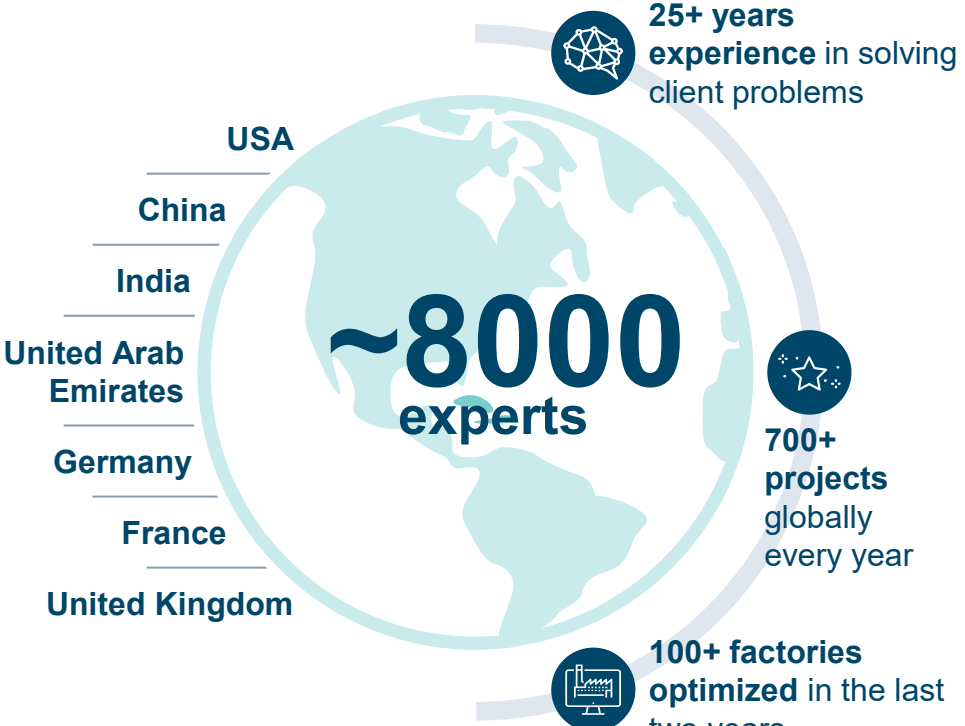
Consulting



Solution Design & Prototyping

Solution Implementation

Interdisciplinary, agile teams to deliver end-to-end consulting and solutions



Source: Siemens Advanta Consulting

SIEMENS

Our team is complementing the ADV Consulting strategy & digital portfolio with Data Science expertise – focusing on Enterprise & Industrial AI applications

Our Team

We employ **40+ data analytics consultants** in **3 countries.**

Our highly diverse team comes from **11 countries** and is fluent in **13 languages**.

Over the past two years, we have successfully **delivered >50** data science & AI **projects**.

Our **interdisciplinary** team has a strong background in academia and a broad data science and data engineering toolset.

Source: Siemens Advanta Consulting

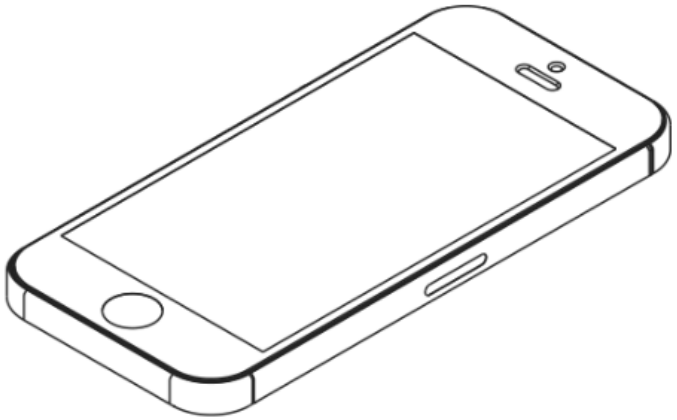
Out of 300.000+ Jobs at Siemens, there is not one that is not impacted by Data Science

Source: Siemens Advanta Consulting

SIEMENS

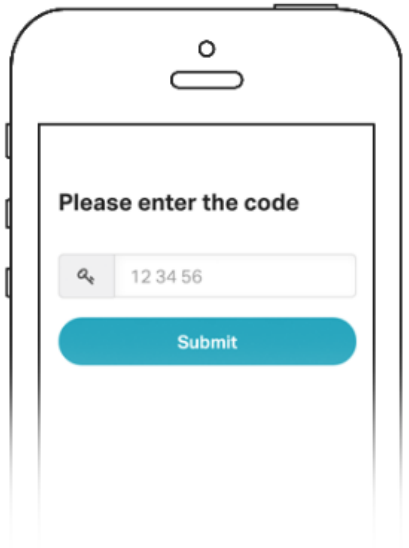
Go to www.menti.com and use the code 4957 5929

What do you estimate is the percentage of Data Science projects that never go into production?



1

2



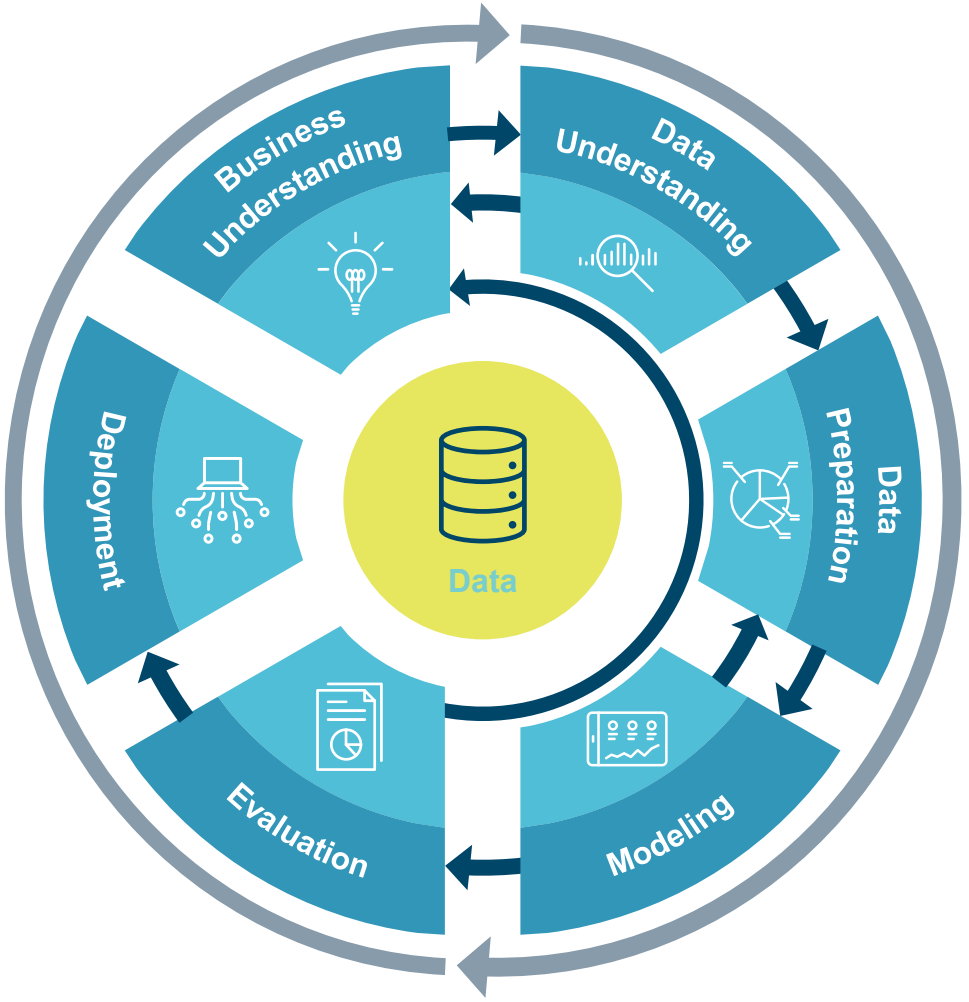
3

SIEMENS

Source: Gartner 2017

SIEMENS

The reasons for Data Science projects to fail are manifold – the Cross Industry Standard Process for Data Mining (CRISP-DM) is one lever to mitigate the risk of failure



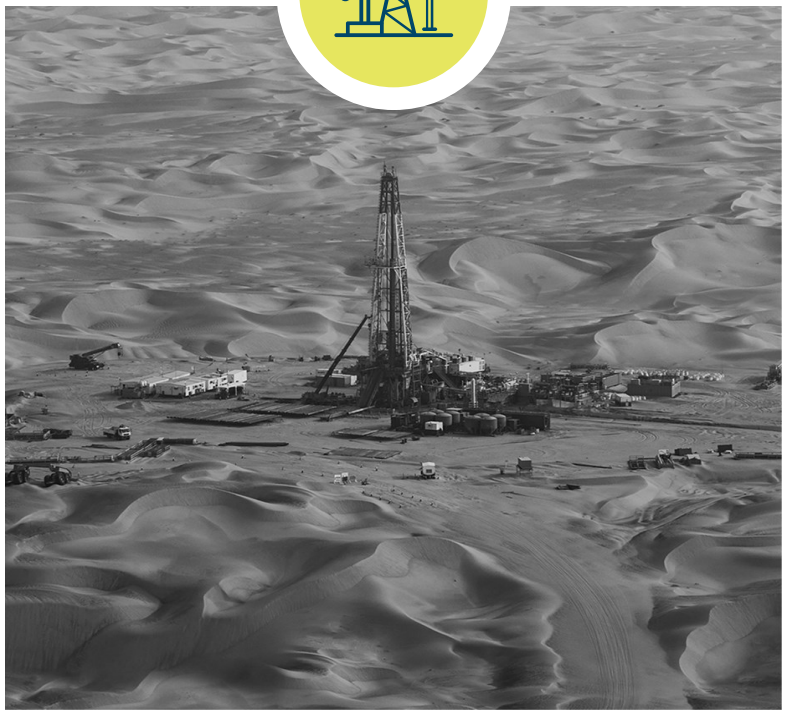
Source: CRISP-DM 1.0 (Chapman et al.)

SIEMENS

The framework is industry agnostic and has helped us in deploying Data Science solutions across a wide range of industries & customers

Healthcare
Remote condition monitoring of medical devices

Oil & Gas
Prediction of sand accumulation in the desert



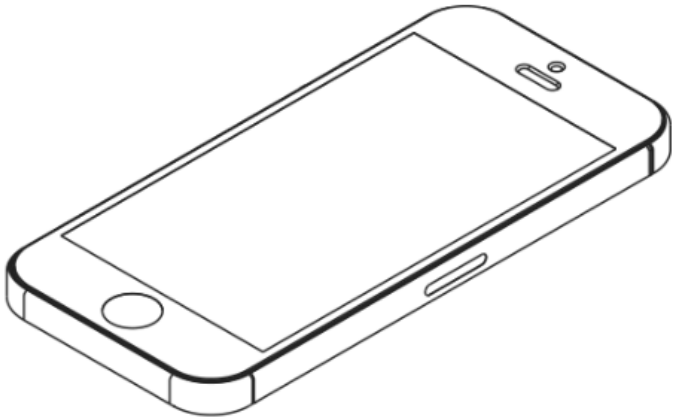
Smart Infrastructure
Decision support for sales representatives to determine the optimal sales price

Source: Siemens Advanta Consulting

SIEMENS

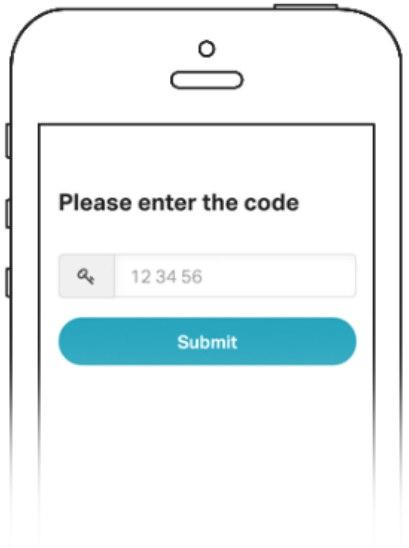
Go to www.menti.com and use the code 4957 5929

In which of the three use cases are you most interested in?



1

2



3

SIEMENS

Healthcare

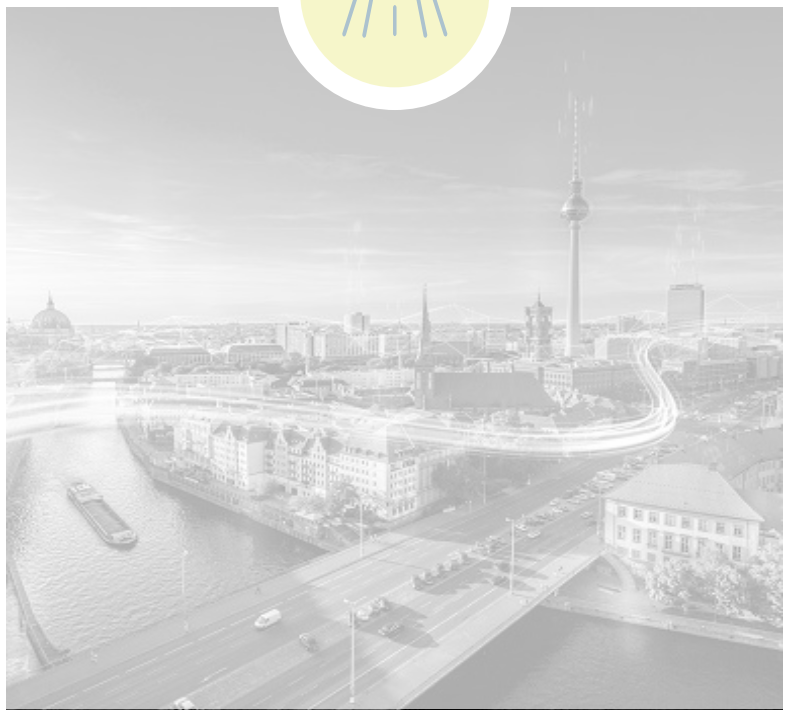
Remote condition monitoring of medical devices

Oil & Gas

Prediction of sand accumulation in the desert

Smart Infrastructure

Decision support for sales representatives to determine the optimal sales price



Source: Siemens Advanta Consulting

SIEMENS

Data-driven solution for condition monitoring to improve patient experience

CLIENT
Multi-national corporation in MedTech

CLIENT CHALLENGES
Lack of data-based insights to improve patient experience and identify business potentials from IoT integration

OUR SOLUTION
Connecting an optimized field device set-up and advanced algorithms to derive condition monitoring solution with underlying business case

CLIENT BENEFITS
Delivery of technical proof-of concept with annual profit potential of several Mio $

Source: Siemens Advanta Consulting

SIEMENS

Home dialysis devices have several pain points due to high returns and undetected faults

Patient

Why is the device so noisy? Am I safe?

Why do so many patients complain about the noise? Is the device really broken? Better return it to be safe!

Technical Support

Failure Analyst

Can't say if the device is really broken, it seems fine.

Why do we have so much buffer capacity, capital expenditures and refurbishment costs?

Business Analyst

Challenges

- High return rates with undetected faults
- 15-18k return/year, but 40% of the returns with no problem found

Approach

- Remotely detect faulty devices
- Classify which part of the device is broken
- Faster fault identification
- Improve patient retention
- Reduce refurbishment and related costs

Lessons Learned

- All stakeholders must be involved at the beginning of a project and their different challenges must be taken seriously.

Source: Siemens Advanta Consulting

SIEMENS

Acoustic sensors could lead to privacy concerns – Vibration is a good alternative that can solve that problem

Challenges

- Collect high-quality data, while taking patients privacy into account (e.g., acoustic data would be inappropriate)

Approach

- Use of vibration data from sensor attached to the device
- Training data collected and labeled by technicians in laboratory setting *vibration file from broken machine? (yes / no)*

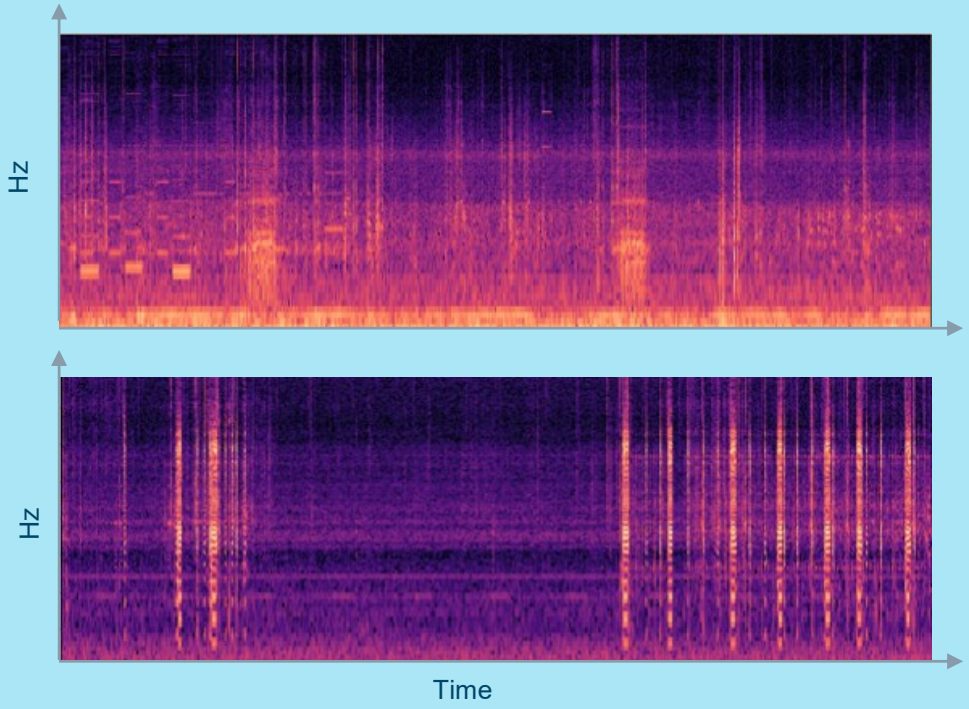
Lessons Learned

- The data quality directly depends on the selection of a suitable sensor and its setup.

Source: Siemens Advanta Consulting

SIEMENS

The vibration data can be used to identify different mechanical components of the machine and their faults



Spectrogram of healthy machine

Spectrogram of faulty machine

Challenges

- Identify which part of the vibration file indicates that the machine is defective
- Deal with limited positive examples

Approach

- External company takes over the data labeling
  focus on individual parts of the vibration file

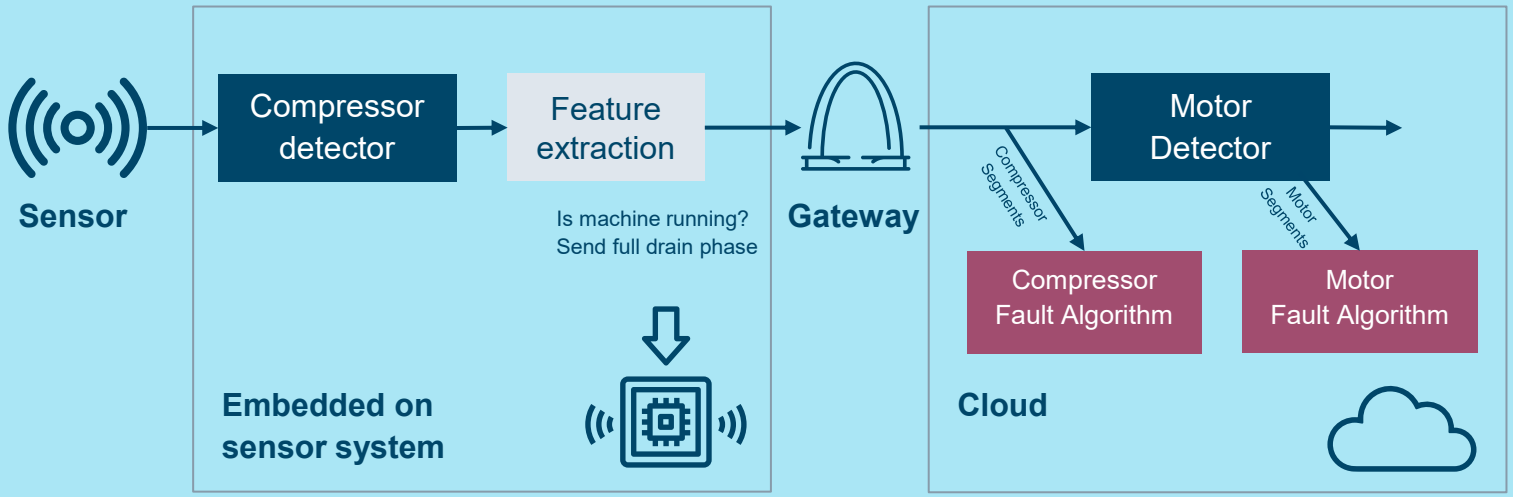
Lessons Learned

- Correct labeling is key and often involves people.
- Data augmentation can help with limited positive examples by creating slightly different copies of the original file.

Source: Siemens Advanta Consulting

SIEMENS

Three models are used to classify the faulty nature of different parts of the machine



Approach

- Model: Random Forest
- 3 models for different parts of the machine

Lessons Learned

- By applying the same model several times and combining their results, a detailed statement about the condition of the device can be made.

Challenges

- Classification problem
- Is machine faulty? (yes / not) > if yes: which part is broken?

Source: Siemens Advanta Consulting

SIEMENS

Data drift in a real environment can lead to the need for adjusting the entire pipeline

Challenges

- Models should also provide good predictions for new data

Approach

- Feasibility tests
- Understand how data is collected in real-world environment and how this affects the model trained with data from a test environment
- Adjustment of the model and pipeline based on the results of the evaluation

Lessons Learned

- An unsatisfactory result does not always have to be caused by the model itself, but can also result from poor data quality or data labeling.
- Often the models need to be adjusted after testing in the real environment and sometimes even the initial target.

Source: Siemens Advanta Consulting

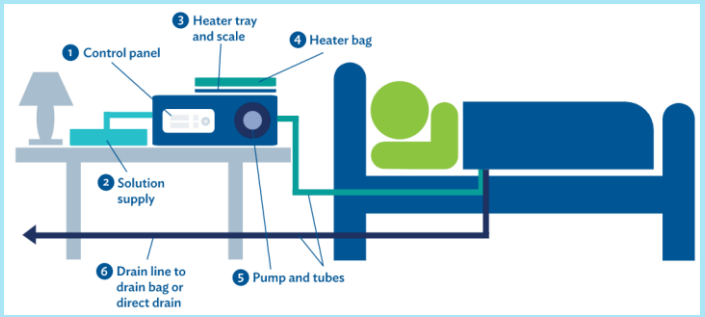
SIEMENS

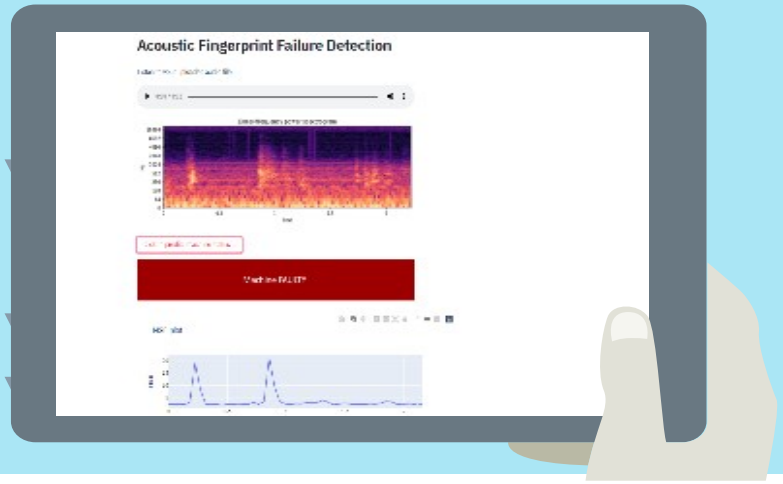
Pains and gains of all stakeholders must be considered in the rollout to create a value proposition

Patient

Technical Support

Failure Analyst





Challenges

- Proper set up of the device at the patient's home
- Overview on the functionality of the devices for the company

Approach

- Demonstrating proper installation of the device with the new sensor in patients' premises
- Condition monitoring app for technical support and failure analyst that provides a full fleet overview and pro-active alert monitoring

Lessons Learned

- The best model is worth nothing if it is not accepted and used. Usability is key!

Source: Siemens Advanta Consulting

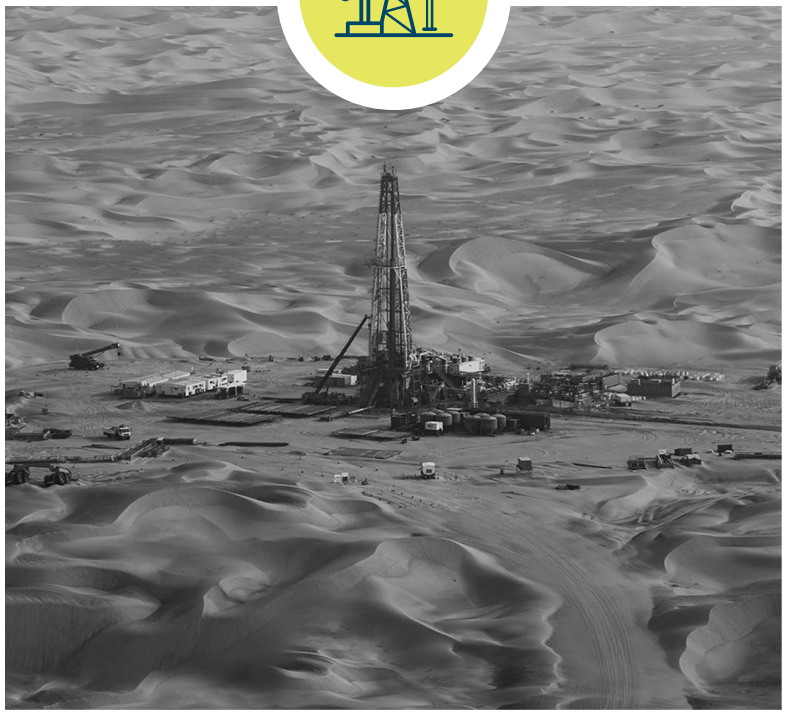
SIEMENS

Healthcare

Remote condition monitoring of medical devices

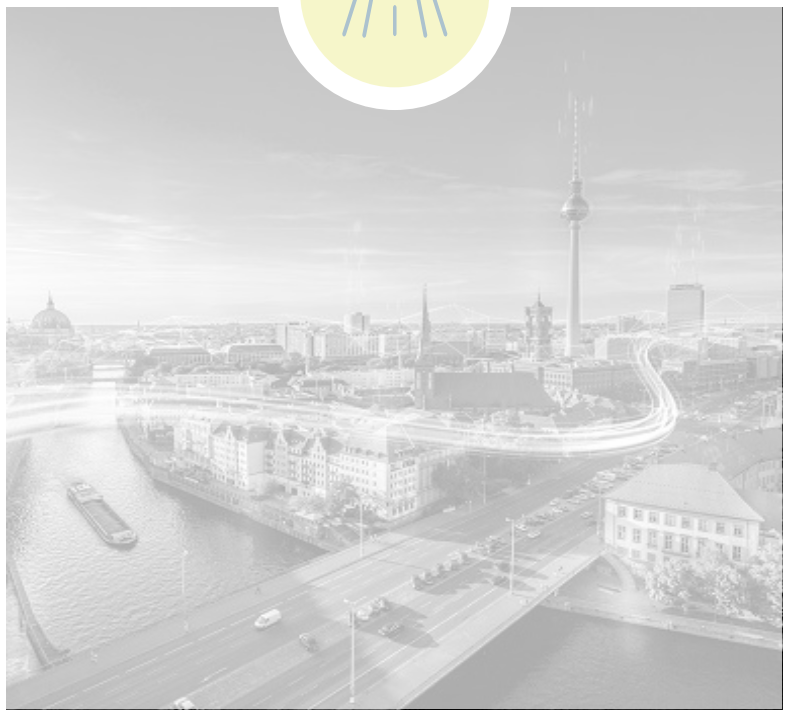
Oil & Gas

Prediction of sand accumulation in the desert



Smart Infrastructure

Decision support for sales representatives to determine the optimal sales price



Source: Siemens Advanta Consulting

SIEMENS

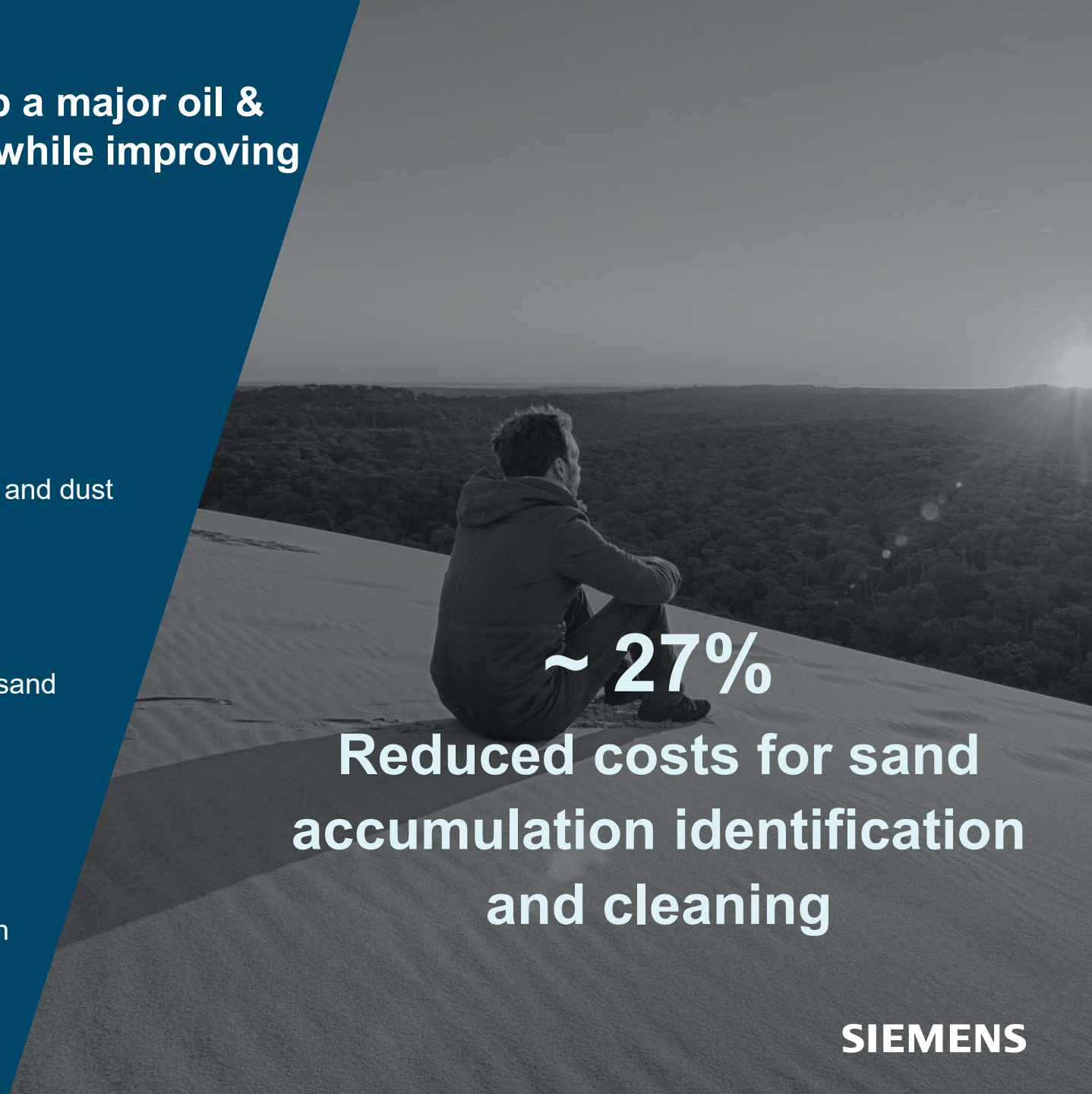
We predicted sand accumulations to help a major oil & gas customer reduce cost of operations while improving plant availability

CLIENT
Leading O&G company in the Middle East

CLIENT CHALLENGES
Disruption of operations due to accumulation of sand and dust on critical infrastructures.

OUR SOLUTION
Leveraging existing data and its potentials to predict sand accumulates based on a machine-learning algorithm.

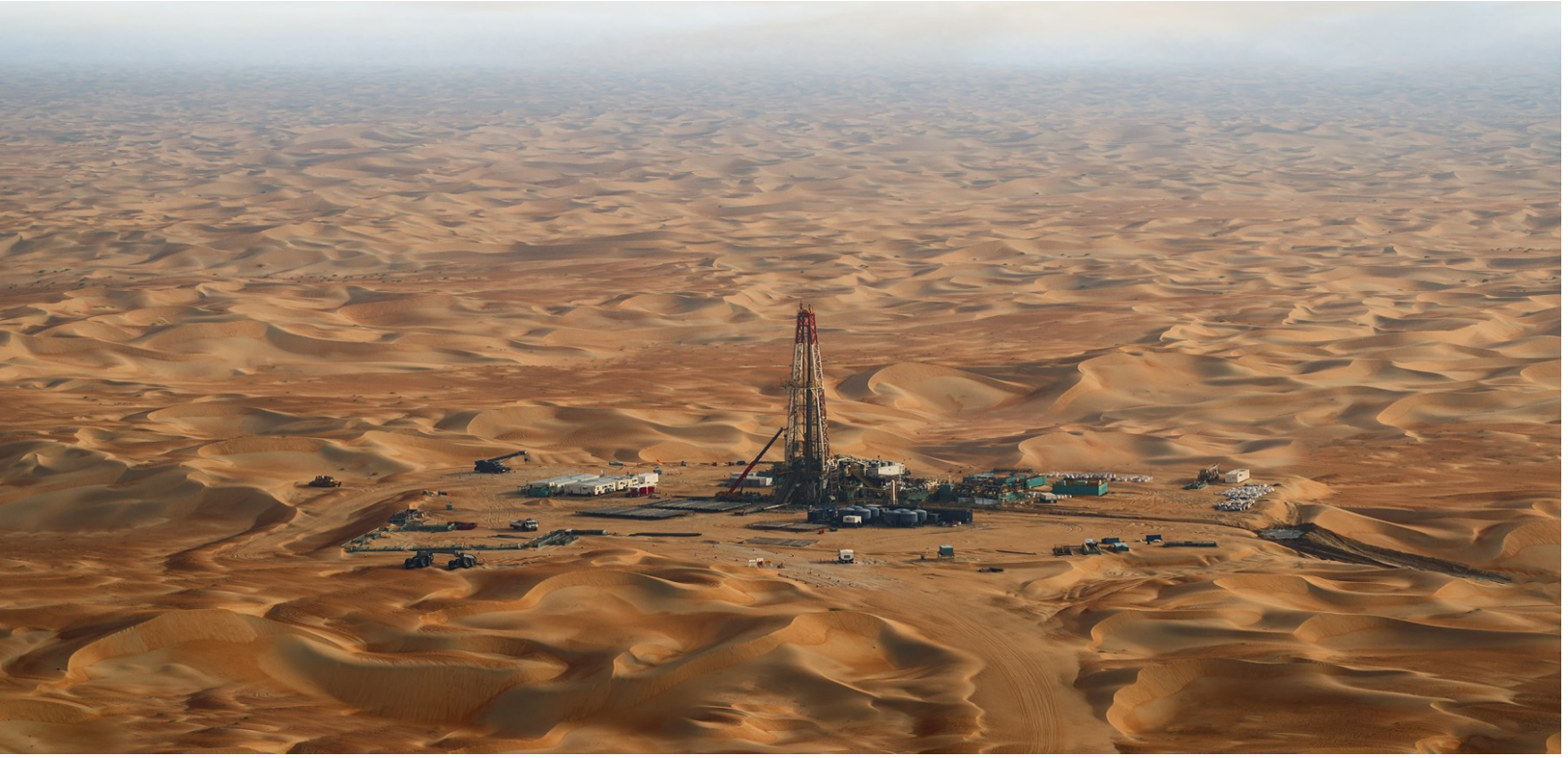
CLIENT BENEFITS
Reduced risk of activity disruptions and cost of sand cleaning, based on prediction of sand accumulation in industrial areas.



Source: Siemens Advanta Consulting

SIEMENS

Sand accumulations and dust block and cover access to roads and disrupt operations



Challenges

- Disruption of operations due to accumulation of sand on critical infrastructure
- Manual checks through patrols required

Approach

- Leveraging existing data and its potentials to predict sand accumulates based on a machine-learning algorithm

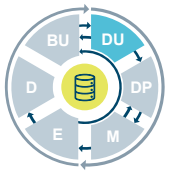
Lessons Learned

- The development of an efficient solution starts from the deep understanding of customer's pain points. First, the problem must be formulated from a business perspective, then it can be determined what data is needed to answer the question.

Source: Siemens Advanta Consulting

SIEMENS

We take five different groups of factors influencing sand movement and accumulation



SOIL	WEATHER	LANDSCAPE	VISUAL	ASSET GENERAL
Sand movement coefficient based on sand type		Elevation	True number of pixels on satellite	
	Humidity	Surrounding elevation	# pixels classified as asset	Asset type
	Temperature		Delta pixels (% to true)	Month
	Wind direction			Compass bearing (asset position)
	Relative wind direction			(to asset position)
	Wind speed			
	Relative surrounding elevation (to wind direction)			
	Type of weather conditions (e.g. rain)			

historical sand cleaning records

Challenges

- Identify factors that have high influence on sand movement

Approach

- We combined available data on clearing activities for every geo-point and trained an algorithm to predict accumulation
- We identified five high influencing factor groups

Lessons Learned

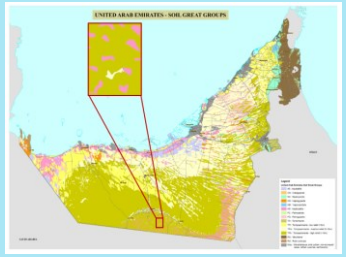
- Various internal and external data is often needed to explain a situation.
- If the data sources are not well selected or some are missing, the model will not produce the intended results.

Source: Siemens Advanta Consulting

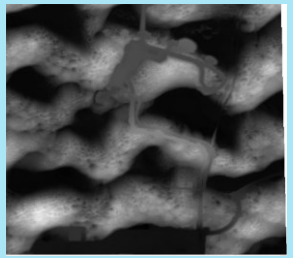
SIEMENS

Pre-processing procedures highly depend on the input data and are key for the final predictive model

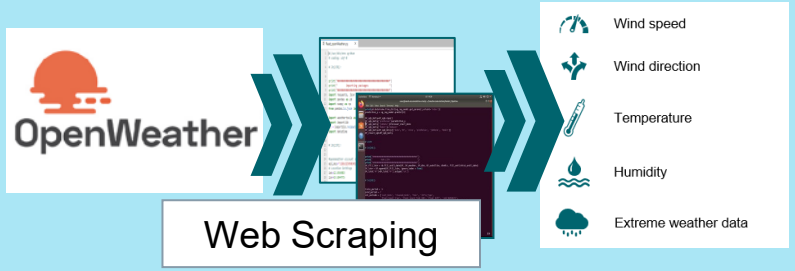
SOIL



LANDSCAPE

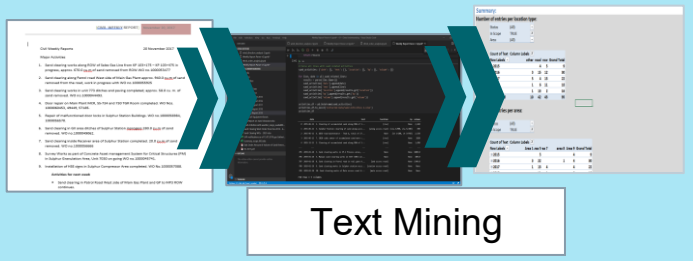


WEATHER



VISUAL

HISTORICAL SAND CLEANING RECORDS



Challenges

- Different data sources and pre-processing procedures for the input factors are required

Approach

- Different input data requires different handling to be used in the final prediction model.
- Text mining, computer vision, web scraping were used for the pre-processing steps.

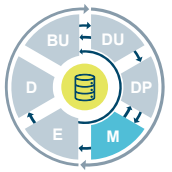
Lessons Learned

- Often data cannot be used directly in the final prediction model.
- Different data types require different preprocessing, for this, ML algorithms are applied, for instance.

Source: Siemens Advanta Consulting

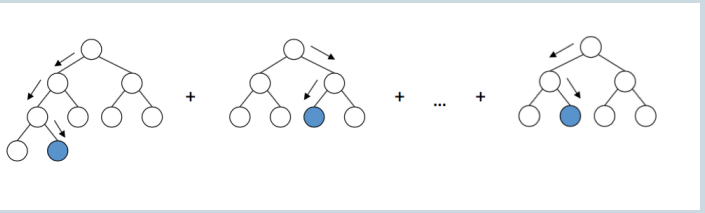
SIEMENS

We used two conceptually different modelling architectures to capture specifics of the data



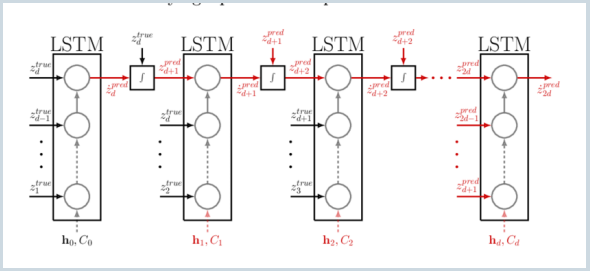
Gradient Boosting

- Partitioning non-linear model
- Explainable on global and local levels
- Tuning with Bayesian grid search
- Consecutive predictions of daily values



LSTM

- Multiplicative non-linear model
- Low-to-no explainability of results
- Random grid search tuning
- Simultaneous prediction for the complete horizon



Approach

- Predict sand movement with two conceptually different modelling architectures: LSTM and XGBoost

Lessons Learned

- A single model is often not sufficient to address the different types of data and derive knowledge from them.

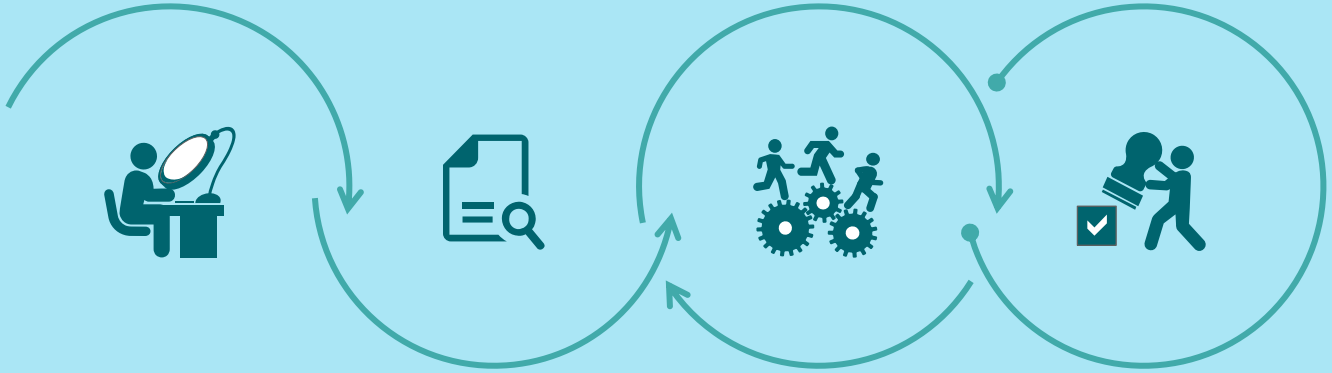
Challenges

- Development of a suitable model that allows good conclusions to be drawn from the various input data

Source: Siemens Advanta Consulting

SIEMENS

User Acceptance Test will be carried out to test scenarios that emulate the user's journey



Identify test scenarios

Identify and prepare test scenarios based on the high-level business process and create test cases with clear steps

Prepare test data

Define and prepare the data required for testing

Execute testing

Run the test cases, identifying bugs (if any) and retesting the bugs once solved

Sign off UAT

Conclude and sign off UAT, ensuring that no critical defects are open and that business objectives are met

Challenges

- Verifying whether the developed solution meets the needs of the end user or whether there is room for improvement

Approach

- Test the model's ability to identify sand accumulation areas and volume
- Test the ability of users to identify the location and amount of accumulated sand

Lessons Learned

- It is not only important for the model to perform well, but also to get the users on board early.
- User acceptance tests are key!

Source: Siemens Advanta Consulting

SIEMENS

The results of data analysis and predictions are delivered to the end users through a tailor-made mobile app

Challenges

- Address the needs of different end users

Approach

- Develop different interfaces for different end users
- Reduce risk of activity disruptions through early detection of sand blocking
- Optimized tracking and scheduling process of sand removal activities

Lessons Learned

- Different interfaces of the final solution help to meet the needs of different end users.

Source: Siemens Advanta Consulting

SIEMENS

Healthcare

Remote condition monitoring of medical devices

Oil & Gas

Prediction of sand accumulation in the desert

Smart Infrastructure

Decision support for sales representatives to determine the optimal sales price

Source: Siemens Advanta Consulting

SIEMENS

Bring the experience of the organization to the fingertips of the sales representatives

CLIENT
Smart Infrastructure

CLIENT CHALLENGES
Lack of consistency and transparency about quotes to customers

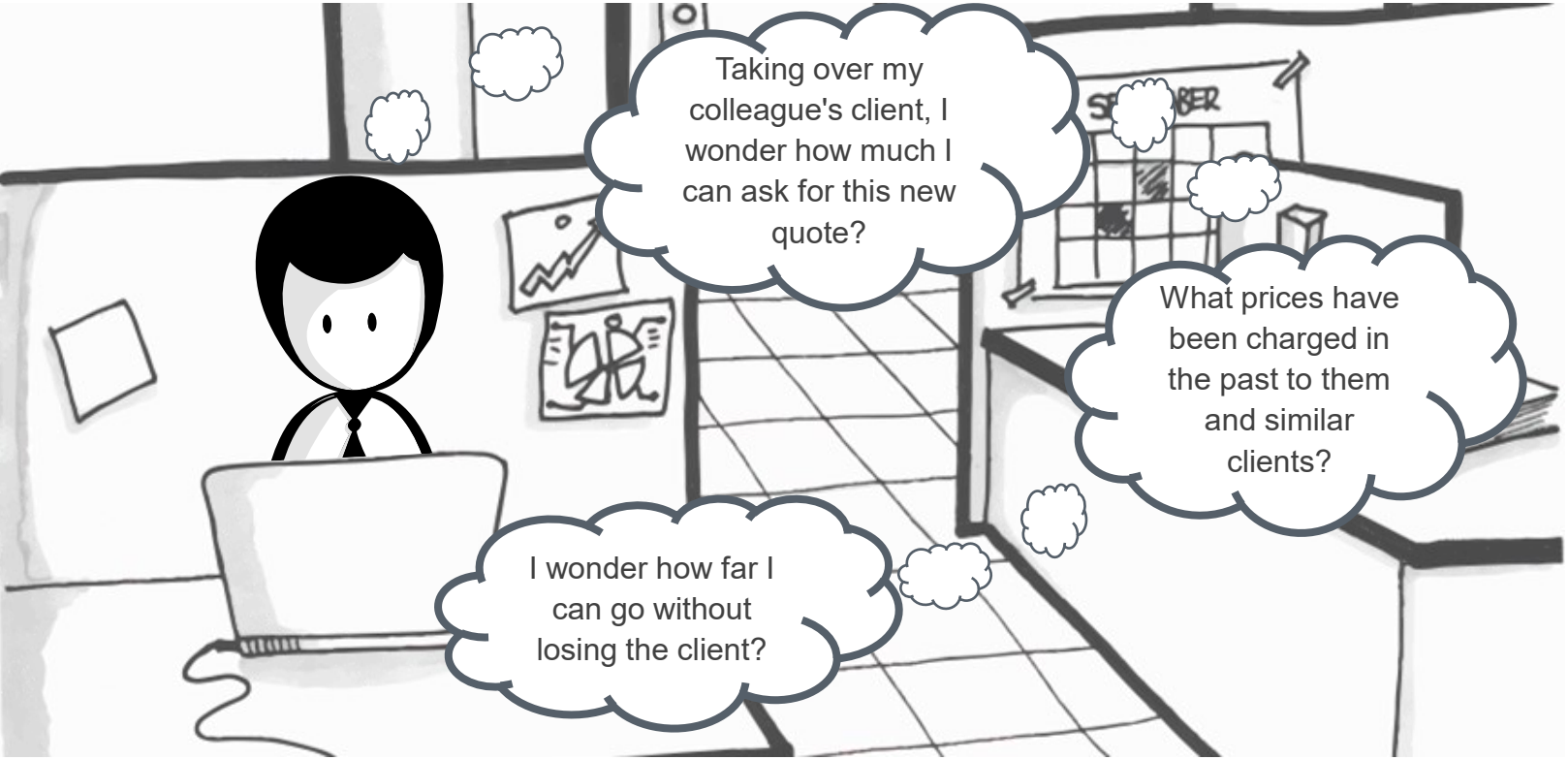
OUR SOLUTION
Automated price recommendation based on the history of sales offers via machine learning

CLIENT BENEFITS
Transparency enhancement and improvement of sales margin

Source: Siemens Advanta Consulting

SIEMENS

Need for decision support in offering prices to new as well as to existing customers based on historic data



Challenges

- Understand the business process and its deviations (per country & business type)
- Identify key contacts to learn about these processes and drive them

Approach

- Get commitment by top management and identify key contacts
- Conduct workshop(s) with business process and governance experts to understand the business process and its deviations

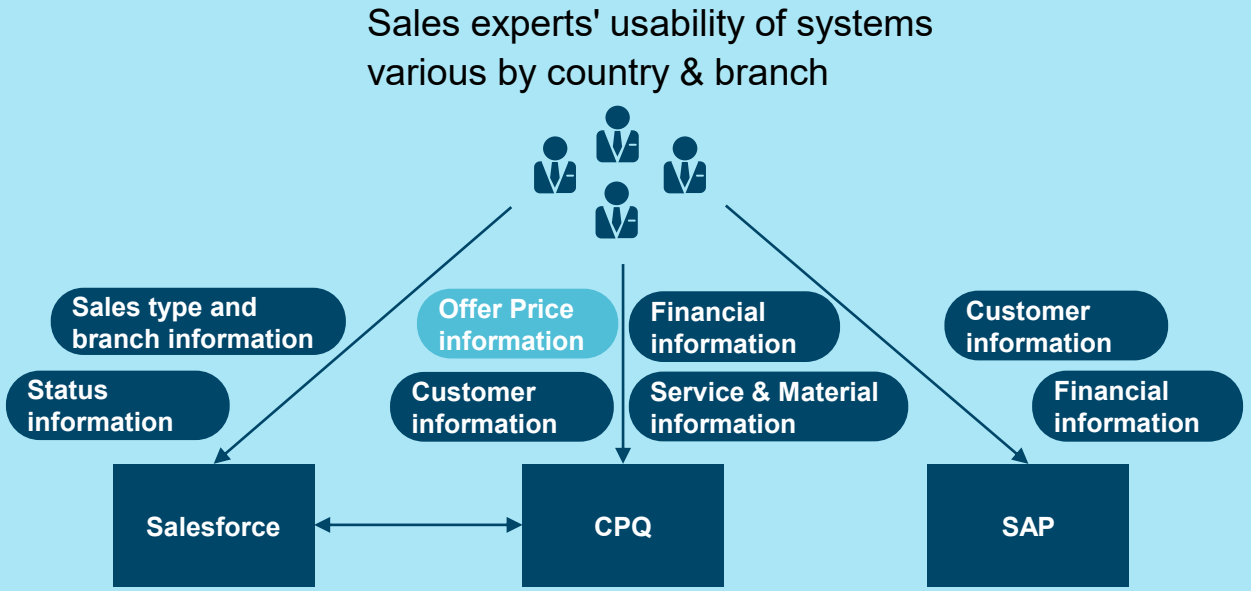
Lessons Learned

- Business processes can vary across country even within the same organization. Pay attention to details and include the major stakeholders.

Source: Siemens Advanta Consulting

SIEMENS

Identifying and connecting the right data sources is one thing, assessing and ensuring high data quality another



Challenges

- Process differences result in data differences
- Identify data experts and get access to data sources

Approach

- Conduct workshop(s) with business / data experts to have a holistic view and identify influencing factors on the price / to data sources
- Compare expert insights with own data analysis

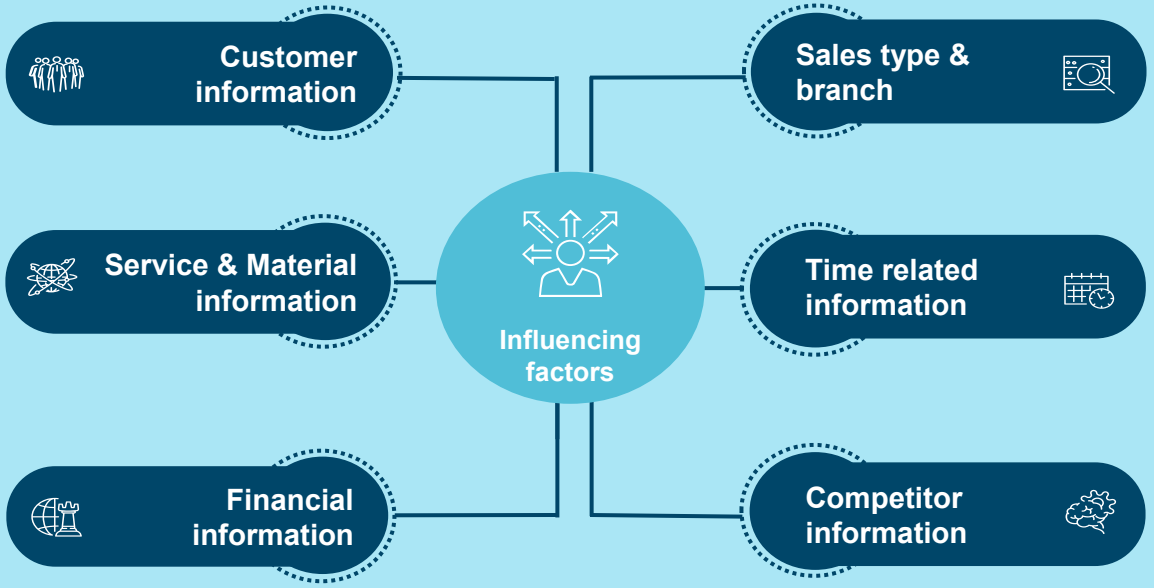
Lessons Learned

- A good business understanding helps significantly to make sense of data. Clarify questions on data early with business to make sure you do have the right understanding.
- Well defined processes are key for consistent data quality.

Source: Siemens Advanta Consulting

SIEMENS

Conduct workshops with business and data experts to identify and map influencing factors



Challenges

- Map identified features to data sources and ensure data quality
- Explainability of features

Approach

- Ensure data quality
- Ensure explainability of features

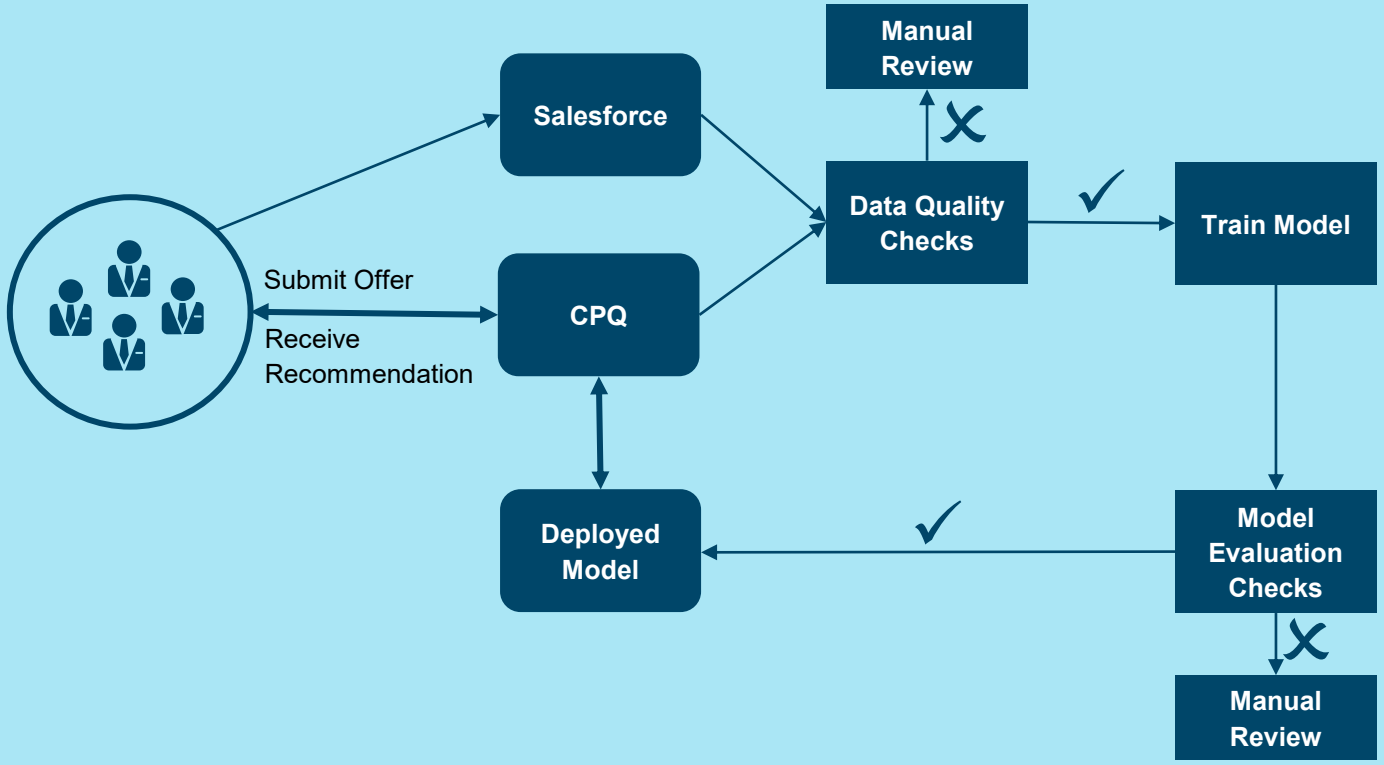
Lessons Learned

- Data preparation has to be done in co-creation with the business and data experts of the customer. Due to data availability & quality, many potential features have to be filtered out.

Source: Siemens Advanta Consulting

SIEMENS

The trained model must be thoroughly evaluated from a technical and business perspective



Approach

- Train model
- Conduct technical evaluation (analyze for systematical model errors)
- Conduct business evaluation (feature importance vs. business understanding)

Challenges

- Standardization vs. individualization
- Consideration of concept drifts (e.g. inflation and Covid)

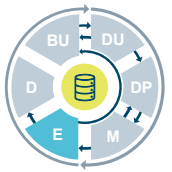
Lessons Learned

- Create models as individualized as necessary, but as standardized as possible! Adjustments for individual countries only if it results in high impact.

Source: Siemens Advanta Consulting

SIEMENS

The calculated impact expectations have been validated via an AB Testing based on real business transactions



Test group
Price Recommender

Control group
No Price Recommender

Challenges
- Model evaluation
- Validation of business impact

Approach
- Conduct AB Testing together with selected countries to prove business impact of AI solution

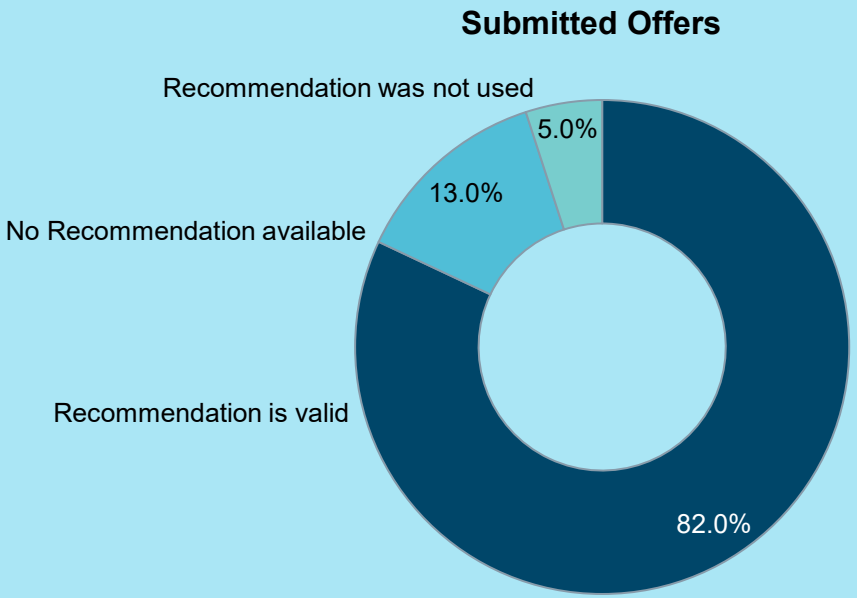
Lessons Learned
- Before the final deployment of a new solution, the expected benefits for the end users must be tested, validated and proven.

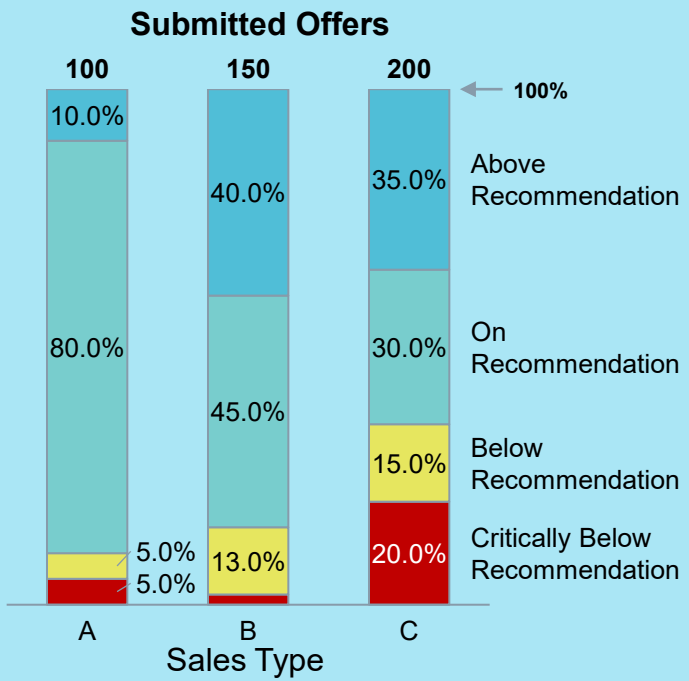
Source: Siemens Advanta Consulting

SIEMENS

Development of a central reporting cockpit to ensure user acceptance and long-term business impact

Price Recommender Dashboard*





* Simplified extracts of parts of the dashboard using dummy data

Challenges

- Integration into existing system
- User centric reporting cockpits

Approach

- Customer co-creation via user stories
- Definition with meaningful KPIs that can track user acceptance and long-term business impact

Lessons Learned

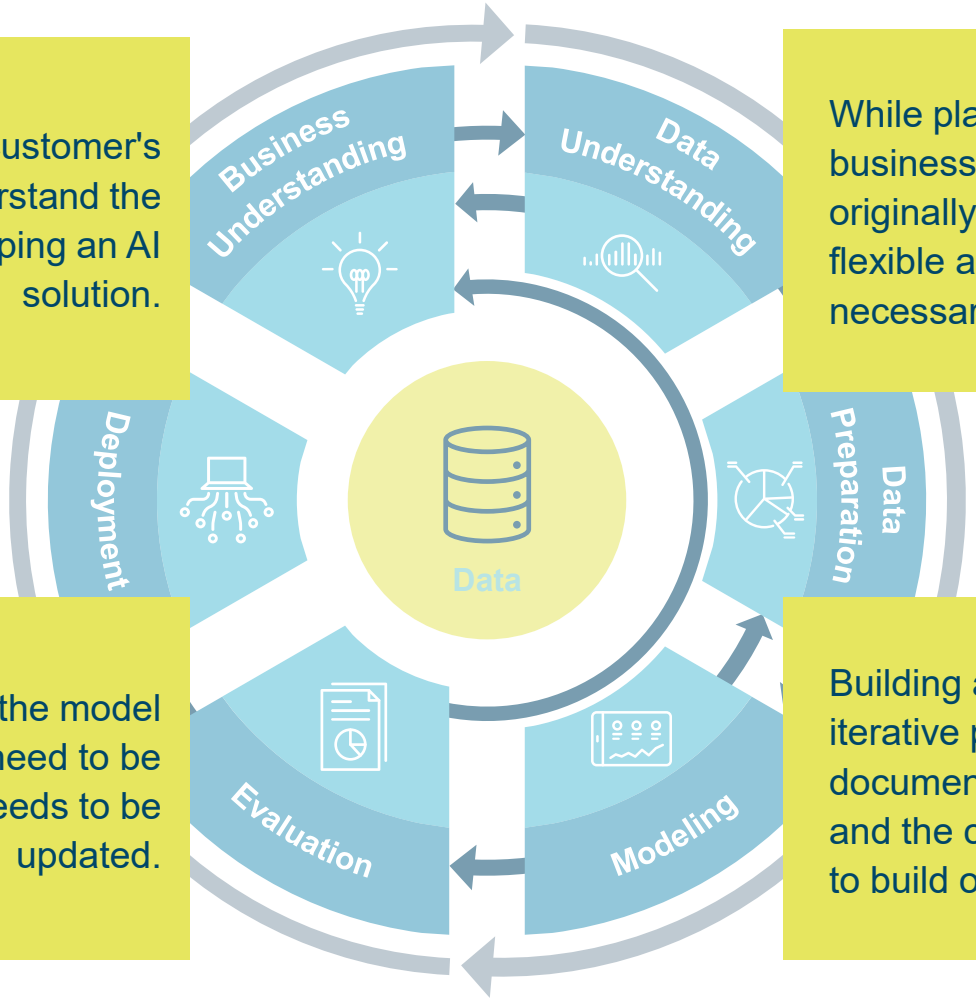
- Only if you can ensure user acceptance and business impact, you create value with the AI solution.

Source: Siemens Advanta Consulting

SIEMENS

The reasons for Data Science projects to fail are manifold – the CRISP-DM framework is one lever to mitigate the risk of failure

Put yourself in your customer's perspective and understand the problem before developing an AI solution.



While planning is important, most business projects do not go as originally planned. It is important to be flexible and adjust the AI strategy if necessary

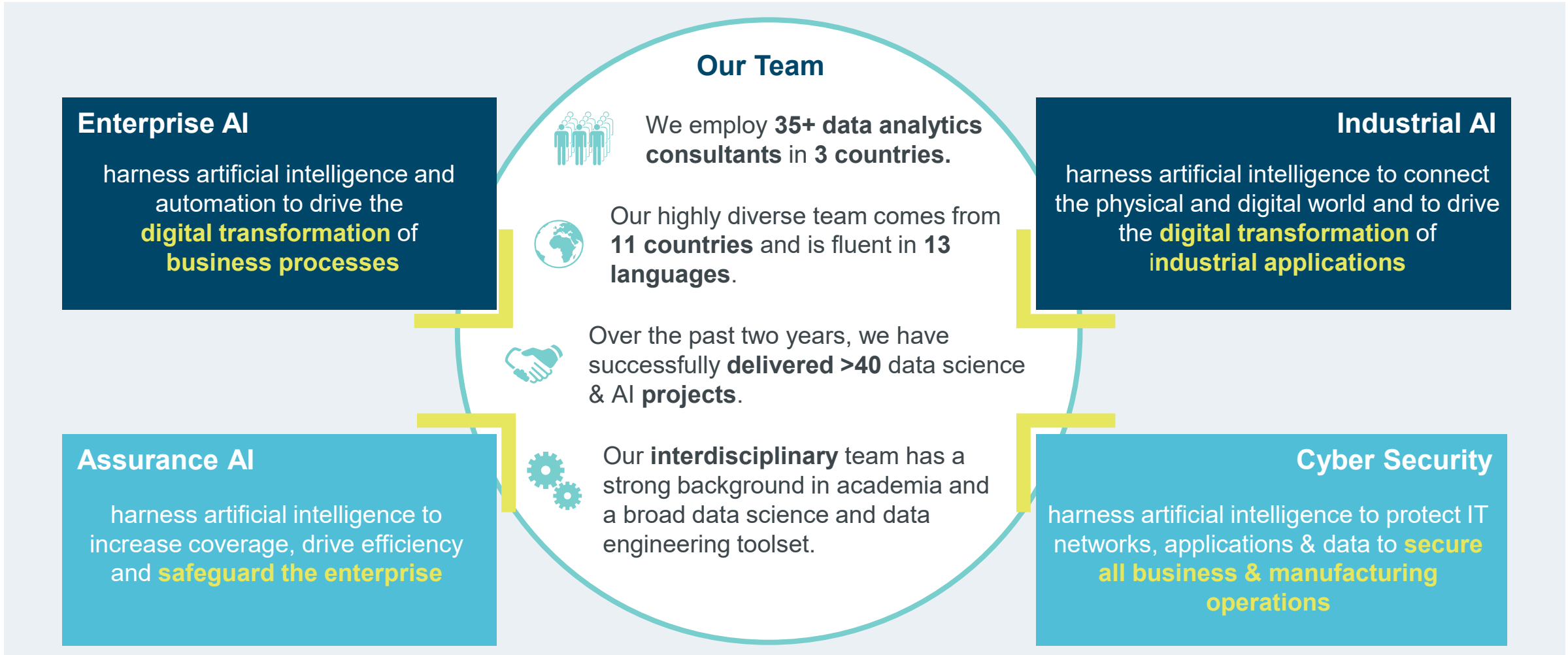
It is important to monitor how the model works with real data. Fixes need to be made or the model needs to be updated.

Building an AI model is usually an iterative process. Systematic documentation of each experiment and the corresponding results helps to build organizational knowledge.

Source: Siemens Advanta Consulting

SIEMENS

Our team offers a unique opportunity to start your career and develop into a leadership role in Data Science within Siemens



Source: Siemens Advanta Consulting

SIEMENS

Q&A

SIEMENS

Thanks for listening

What's next?

- Apply on siemens.com/smc
- Come and talk to me afterwards
- Email us:
  - recruiting.advanta-consulting@siemens.com
  - Jochen.Gross@siemens.com
- Find us on **LinkedIn**

Not done studying?
Still, get in touch with us!

Siemens offers …

- Bachelor & Master theses,
  Working student positions → **Karin Bründl**
- PhD/PostDoc → **Fabian Rhein**
- Internship, Consultative expert talk, CV check,
  8h-Monitoring → **Fabian Rhein** & **Karin Bründl**

Siemens Advanta offers …

- Master theses, Internships,
  Working student positions → **Edward Graf**