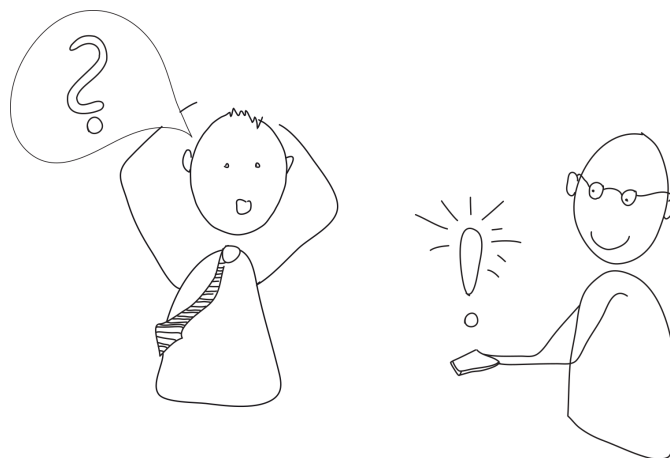


# A Hitchhiker's Guide to a Data Science Use Case

## Introduction

The first step towards a successful Data Science project is understanding your use case. This guide is the result of over 12 years of field experience in Data Science that taught us one lesson: There are no simple answers! Hence this guide focuses on providing the questions which usually arise during the course of a Data Science project. Many of them arise only after one is stuck at a dead end, but following the guide you can avoid this now. Take your time to think about the following questions at the very beginning or better even before you start a project. Answering them thoroughly will pay back soon, as this approach not only helps to avoid costly pitfalls, but will also help to sharpen your understanding of the use case and identify areas where you miss expertise. Together this will help you to successfully implement a data science project and help to establish data science in your company!



Answering the questions will approximately take 10-15 minutes.

## Old World Computing - Establishing the Future

Old World Computing is an expert consulting company in the area of Data Science, focusing on predictive analytics and big data. Our team can rely on experience gained from more than twelve years in the field of applying predictive analytics using RapidMiner, a world leading Data Science Platform. Every single consultant is certified by RapidMiner Inc.

Based on this experience, we devised our unique Collaborative-CRISP approach. Following this process, we support our customers during all phases of establishing Data Science in a company: From laying out the strategy together, through training courses and knowledge transfer projects to train internal resources, up to delivering turn-key solutions and long-term expert support to secure and improve what has been achieved.

We combine our field experience with a back office development team whose purpose is to support our consultants and our customers' data scientists in their projects. If a project requires a special functionality that the RapidMiner analytics platform does not support, our agile development team adds this feature. This way we do not only ensure the highest efficiency of our consultants during customer projects, but the customers' data science team also benefits from the real-world proven functionality we add.

Our expertise has made Old World Computing a listed partner of RapidMiner Inc. and the official RapidMiner training partner for Central Europe, delivering every official training in the DACH region.

Our customers mainly are medium to large enterprises in the automotive, telecommunication or energy sector and our projects mostly focus on Industry 4.0, predictive maintenance and improving customer satisfaction.

**Now you know about us. Let's see what we can do for you on the following pages!**

**If you need additional information, please visit us at [www.oldworldcomputing.com](http://www.oldworldcomputing.com) or write to: [contact@oldworldcomputing.com](mailto:contact@oldworldcomputing.com).**

# Use Case Definition

## Type

- research project
- proof of concept
- business case

## Deployment

- in scope
- out of scope
- what is that?

# Organizational Setup

## The project is initiated by

- single person
- team
- department
- management

## Experience of the initiator

- first project
- first project to deploy
- business as usual

## The Companies Data Science Team

- is involved
- just focuses on aspects
- only stores data
- no cooperation yet

- not existent

## The companies Central IT

- is supporting
- is involved
- doesn't know about the project

## Is the Budget fixed already?

- yes
- no

## Is the Technical infrastructure fixed already?

- completely
- partly
- green field

# Targeted Audience

Which group of people will use the project results?

---

---

---

---

---

## Charaterize this group

- domain experts
- management
- worker
- machines
- other: \_\_\_\_\_

## Importance of the results to this peoples work

- business critical
- decision support
- nice to have
- looks good
- distracting

## Degree of automation aimed for

- full
- decision support
- feedback loop with human
- insight generation

## Results will be presented as

- WebServices
- interactive web apps
- interactive reports
- excel reports
- Powerpoints

**Which business problem of the user group is addressed by the project?**

---

---

---

---

---

**Does the project require domain knowledge from the user group?**

- yes
- no

## **Impact**

**How are the results helping the user group in their daily business?**

---

---

---

---

---

### **Impacts of Results**

- Job Loss due to automation
- persisting experience
- eases daily work
- creates additional noise
- creates additional work

**How are the results integrated in user's daily work?**

---

---

---

---

---

**Value Generation - savings per year [\$]**

- inestimable
- > 10.000
- > 100.000
- > 1.000.000
- business critical

**Impact on company**

---

---

---

---

---

**Impact on society and humankind as a whole**

---

---

---

---

---

---

## Data

### Data source

- owned by me or my team
- public data
- 3rd party data
- another department's data
- comes from the companies data lake

### Data is generated by

- Sensors
- machine generated
- human interaction recorded by machines
- manually created

### Data is available as

- raw data
- aggregated data
- reports

### Ground Truth

- available as we predict future events
- available as generated manually
- needs to be manually generated
- ground truth?

### Historical Data is available for

- > 5 years
- > 1 years

## Infrastructure

### Avialable Storages for Data

- Hadoop
- NoSQL Databases
- SQL Databases
- Excel

- > 3months
- just started to save
- no historical data available

### Relevant Data set size (total)

- > 1 TB
- > 100GB
- > 1GB
- smaller

### Biggest data set size for one aspect

E.g. if your data is about machine data, how big is the data set for just one machine as the unit you want to analyze?

- > 1 TB
- > 100GB
- > 1GB
- small

### Data is

(multiple are possible)

- tabular
- free texts
- images
- sounds
- other: \_\_\_\_\_

- other: \_\_\_\_\_

### Available Experience Level with Hadoop

- good
- beginner
- none

**Available Experience Level with NoSQL**

- good
- beginner
- none

**Available Experience Level with SQL**

- good
- beginner
- none

**Preferred Analytical environment**

- Graphical
- Code Based

**Analytical Environment already fixed?**

- yes
- no

**Available Experience Level with Analytical Environment**

- good

- beginner
- none

**Available Hardware**

- anything
- cloud based
- virtual standard server
- office PCs
- no dedicated hardware

**Deployment Requirements**

**(multiple are possible)**

- central
- world-wide distributed
- near real-time deployment (sub second reaction time)
- online deployment (seconds reaction time)
- offline deployment