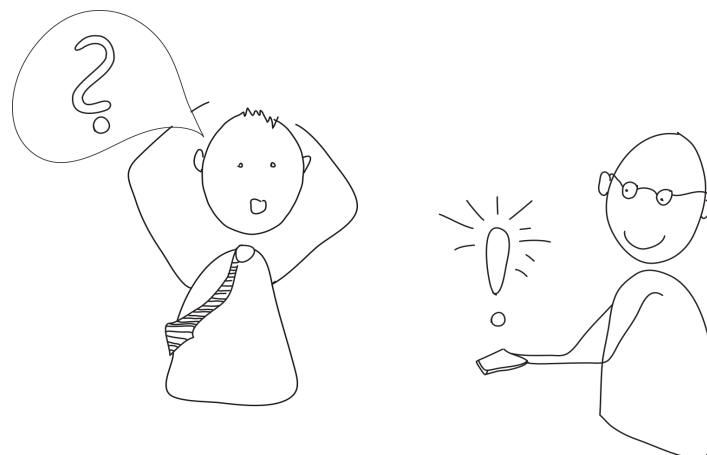


Hitchhiker's Guide to a Data Science Use Case

Einleitung

Der erste Schritt in Richtung eines erfolgreichen Data Science Projekts ist es, den Use Case zu verstehen. Dieser Guide ist das Resultat von über zwölf Jahren Erfahrung im Bereich Data Science, in denen wir eins gelernt haben: Es gibt keine einfache Standardlösung! Der Guide gibt Ihnen deshalb keine Antworten, sondern navigiert Sie mit den Fragen, die normalerweise erst auftauchen, wenn Sie sich schon in einer Sackgasse verlaufen haben, durch Ihr Data Science Projekt. Geben Sie sich Zeit, über die folgenden Fragen zu Projektbeginn, oder besser noch davor, nachzudenken. Die Fragen gründlich zu beantworten, wird sich schnell auszahlen: Sie können so nicht nur teure Fehlinvestitionen vermeiden, sondern auch ein tieferes Verständnis Ihres Use Cases bekommen und einsehen, in welchen Bereichen Sie noch Expertise benötigen. Dieser Guide wird Ihnen helfen, Ihr Projekt erfolgreich durchzuführen und Data Science in Ihrem Unternehmen zu etablieren. Die Beantwortung der Fragen wird etwa 10 bis 15 Minuten in Anspruch nehmen. Nutzen Sie die Ergebnisse für sich oder kontaktieren Sie uns - auf Basis Ihrer Antworten können wir Sie individuell beraten und gemeinsam die nächsten Schritte in Richtung Ihres Data Science Projekts planen.



Use Case Definition

Typ

- Forschungsprojekt
- Proof of Concept
- Businessscenario

Deployment

- In Scope
- Out of Scope
- Was ist das?

Organisatorische Voraussetzungen

Das Projekt wird initiiert von

- einer einzelnen Person,
- einem Team,
- einer Abteilung oder
- dem Management.

Erfahrung des Initiators

- Erstes Projekt
- Erstes Projekt in produktiver Verwendung
- Ein Projekt unter vielen

Das Data Science Team des Unternehmens

- ist beteiligt,
- ist nur für bestimmte Aspekte zuständig,
- speichert nur die Daten,
- arbeitet bisher nicht mit oder
- gibt es noch nicht.

Die zentrale IT des Unternehmens

- unterstützt,
- ist aktiv beteiligt oder
- weiß nichts von dem Projekt.

Ist das Budget bereits festgelegt?

- Ja
- Nein

Ist die technische Infrastruktur bereits festgelegt?

- Komplett
- Teilweise
- Noch gar nicht

Zielgruppe

Welche Gruppe wird die Projektergebnisse nutzen?

Beschreiben Sie diese Gruppe

- Fachexperten
- Manager
- Arbeiter
- Maschinen
- Andere: _____

Bedeutsamkeit der Ergebnisse für die Arbeit dieser Gruppe

- Businessrelevant
- Entscheidungsunterstützung
- Nice to have
- Sieht gut aus
- Ablenkend



Die Ergebnisse sollen in der folgenden Form dargestellt und nutzbar gemacht werden:

- Webservices
- Interaktive Webanwendungen
- Interaktive Reports
- Excel-Reports
- Powerpoints

Angestrebter Grad der Automatisierung

- Vollständig
- Entscheidungsunterstützung
- Feedback-Schleife mit Menschen
- Insight-Generierung

Welches Problem der Endbenutzergruppe wird mit dem Projekt behandelt?

Benötigt das Projekt Fachwissen der Benutzergruppe, um zu gelingen?

- Ja
- Nein

Auswirkungen

Wie unterstützen die Ergebnisse des Projekts die Benutzergruppe in ihren täglichen Aufgaben?

Auswirkungen der Ergebnisse

- Jobverlust wegen Automatisierung
- Erzeugt zusätzliche Arbeit
- Erzeugt störende Faktoren
- Vereinfacht die tägliche Arbeit

Wie sind die Ergebnisse in den Arbeitsalltag der Endbenutzergruppe integriert?

Wertschöpfung - Ersparnisse pro Jahr [€]

- Unschätzbar
- > 10.000
- > 100.000
- > 1.000.000
- Businesskritisch

Auswirkungen auf das Unternehmen

Auswirkungen auf die Gesellschaft und die Menschheit als Ganzes

Daten

Datenquelle/n

- Ich/Jemand aus meinem Team besitzt die Daten
- Öffentliche Daten
- Daten Dritter
- Daten einer anderen Abteilung
- Data Lake des Unternehmens

Die Daten werden/wurden generiert von

- Sensoren
- Maschinen
- von Maschinen aufgezeichneten menschlichen Interaktionen
- manuell erstellt

Daten liegen vor als

- Rohdaten
- Aggregierte Daten
- Reports

Ground Truth

- Verfügbar, da zukünftige Events vorausgesagt werden sollen
- Verfügbar, da manuell erstellt
- Muss noch manuell erstellt werden
- Was ist das?

Historische Daten liegen vor für

- > 5 Jahre
- > 1 Jahr
- > 3 Monate
- Gerade erst angefangen, Daten zu sammeln
- Es gibt noch keine historischen Daten

Größe relevanter Daten insgesamt

- > 1 TB
- > 100GB
- > 1GB
- Kleiner

Größter Datensatz für einzelnen Aspekt

Wenn Ihr Datensatz beispielsweise Informationen über viele Maschinen enthält, wie groß ist der Teil der Daten im Bezug auf nur eine Maschine, als zu analysierende Einheit?

- > 1 TB
- > 100GB
- > 1GB
- Kleiner

Infrastruktur

Verfügbare Speicherformen für Daten:

- Hadoop
- NoSQL Databases
- SQL Databases
- Excel
- Andere: _____

Erfahrung mit Hadoop:

- Gut
- Anfänger
- Keine

Erfahrung mit NoSQL:

- Gut
- Anfänger
- Keine

Erfahrung mit SQL:

- Gut
- Anfänger
- Keine

Bevorzugte Analyseplattform:

- Grafisch
- Codebasiert

Die Daten sind

- Tabellen
- Texte
- Bilder
- Sounddateien
- Andere: _____

Ist die Analyseplattform bereits festgelegt?

- Ja
- Nein

Erfahrung mit bestehender Analyseplattform:

- Gut
- Anfänger
- Keine

Verfügbare Hardware:

- Alles vorhanden
- Cloud-basiert
- Virtual Standard Server
- Office PCs
- Keine bestimmte Hardware

Anforderungen an das Deployment

(die produktive Verwendung der Ergebnisse)

- Zentral
- Weltweit verteilt
- Nahezu Echtzeit-Deployment
(Millisekunden bis zur Antwort)
- Online-Deployment (Sekunden bis zur Antwort)
- Offline-Deployment