CRISP DM Summary:

**Was ist das Problem, das wir haben könnten?**

Was wurde gemacht und warum?

Zentraler Kern des wissenschaftlichen Arbeitsansatzes ist die Reproduzierbarkeit &   
Verifizierbarkeit.

* Was kann schief gehen?   
  <https://www.newyorker.com/news/john-cassidy/the-reinhart-and-rogoff-controversy-a-summing-up>

Das sollte recht einfach sein -> Code und Datenstatz runterladen, kompilieren und

laufen lassen.   
Warum ist es nicht so leicht??

Viel passiert im Doing! Man beginnt, testet Hypothesen, ändert variablen, ….

Bild von Bruegel

**Was ist die Lösung?**

* **Strukturierte, Standadisierte Prozesse**
* **Dokumentation & Kommunikation**

Fundamental Perspective: Data Science vs. Alchemy vs. Art

Fokus muss hin zu Reproduzierbarkeit und damit Verifizierbarkeit (können die Ergebnisse überhaupt überprüft werden? )

* Über Strukturierte Prozesse & Dokumentationen

Um Vertrauen, Effizienz und Korrektheit zu gewährleisten

**Und wie setzt man das um?**

Mit, \*Trommelwirbelllllll\*

CRISP DM, das steht für   
CRoss Industry Standard Process for Data Mining

Ursprünglich aus Team von:

* Daimler-Chrysler
* SPSS
* NCR

Laut KD Nuggets-Umfragen aus den Jahren 2007 und 2014 ist CRISP-DM die am weitesten verbreitete Methode für Data-Science- und Analyseprojekte: Über 40 % der Befragten gaben an, sie zu verwenden.

Auch laut wissenschaftlicher Literatur der Industrie Standard:   
https://www.sciencedirect.com/science/article/pii/S1877050921002416

A close-up of a document

Description automatically generated

Anders als andere Prozessmodelle wie z.B.:

* Fayyad’s KDD
* SEMMA (von SAS)

Liegt der Fokus nicht nur auf die zentralen Data Science Tasks an sich, sondern beinhaltet den ganzen Projekt Prozess also auch Themen wie Deployment & Business Understanding:

A diagram of a data processing process

Description automatically generated

Phasen:

ISO 22989 als Ergänzung erwähnen!! A table of data analysis

Description automatically generated with medium confidence

Tasks sind Fett markiert, Ergebnisse sind in kursiv markiert.

1. Business Understanding
   1. Business Objectives
      1. Ergebnis: Verständnis des Hauptziels des Kunden aus geschäftlicher Sicht
      2. Aktivitiäten
         1. Informelles Beschreiben des zu lösenden Problems
         2. Spezifizieren des erwarteten Nutzens aus geschäftlicher Sicht
         3. Zusatz durch EU AI Act: Risikoklassifikation vornehmen  
              
            A diagram of a triangle with different colored circles and black text

            Description automatically generated
   2. Business Success Criteria
      1. Ergebnis:
         1. Erfolgskriterien mit folgenden Eigenschaften:
            1. Spezifisch
            2. Messbar
            3. Allgemein
            4. Subjektiv (Angabe der messenden Person)
      2. Aktivitäten:
         1. Bestimmen der Erfolgskriterien
         2. Festlegen der Person(en) die die Erfolgskriterien bewerten
         3. Achtung: Jedes Erfolgskriterium sollte sich mindestens auf eine Business Objective beziehen ☺