



Business Understanding für Data Science

Stefanie Scholz

Kontakt



Prof. Dr. Stefanie Scholz

Professorin für Sozialwirtschaft

Marketing – Advanced Analytics – AI

stefanie.scholz@srh.de

SRH Wilhelm Löhe Hochschule | Stefanie Scholz



Agenda

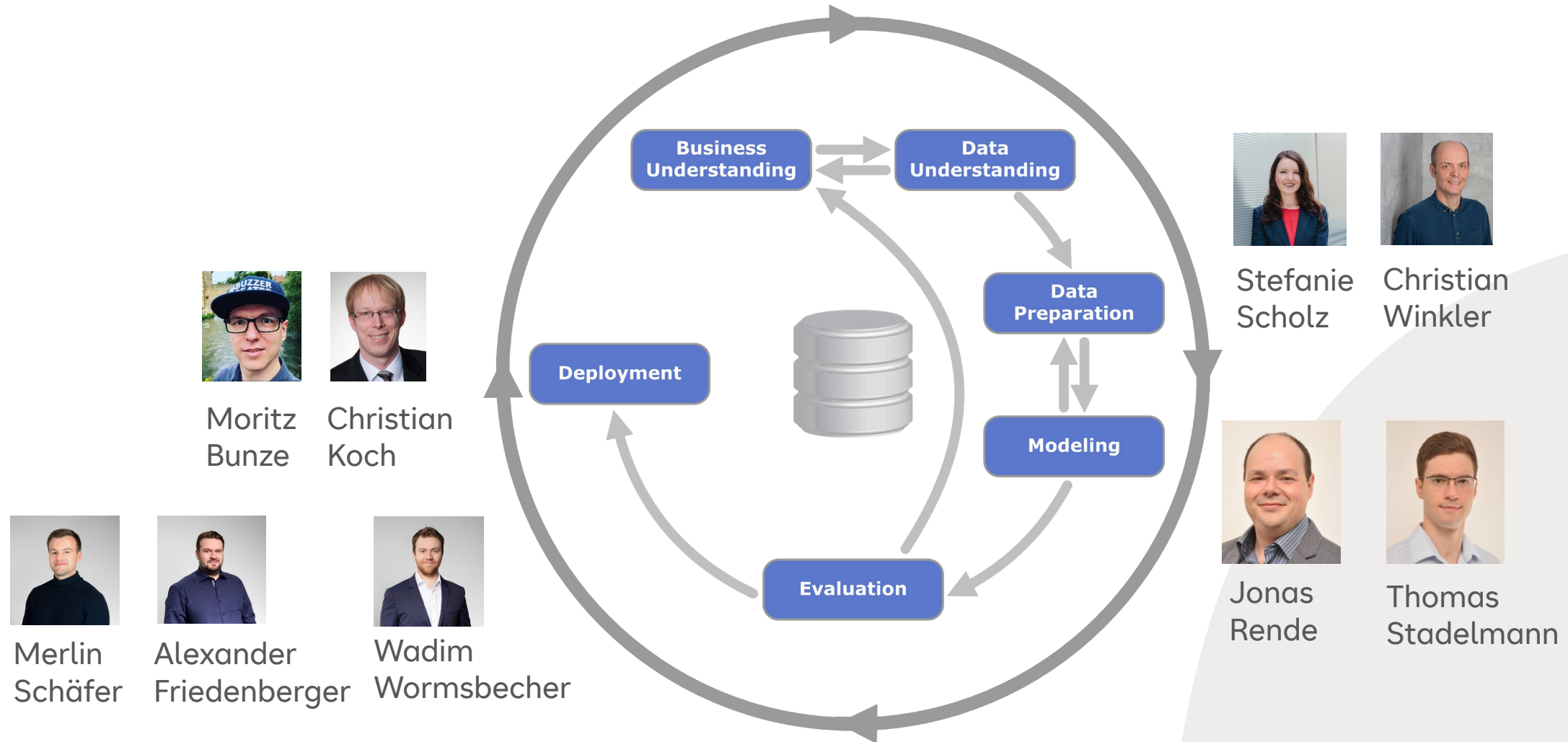


- 1) CRISP-DM als Grundlage für Business & Data Understanding
- 2) Use Case
- 3) Soziale Netzwerke
- 4) Projektablauf einer UGC-Analyse
- 5) Unser Beispiel: Subreddit „r/technology“

CRISP-DM

01

CRISP-DM → Cross Industry Standard Process for Data Mining



https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

CRISP-DM → Cross Industry Standard Process for Data Mining



1) Business Understanding: Anwendungsfall diskutieren und Geschäftsziel verstehen

→ Verständnis der Ziele und Anforderungen aus der Business Perspektive & Transfer in die „Data Science-Welt“

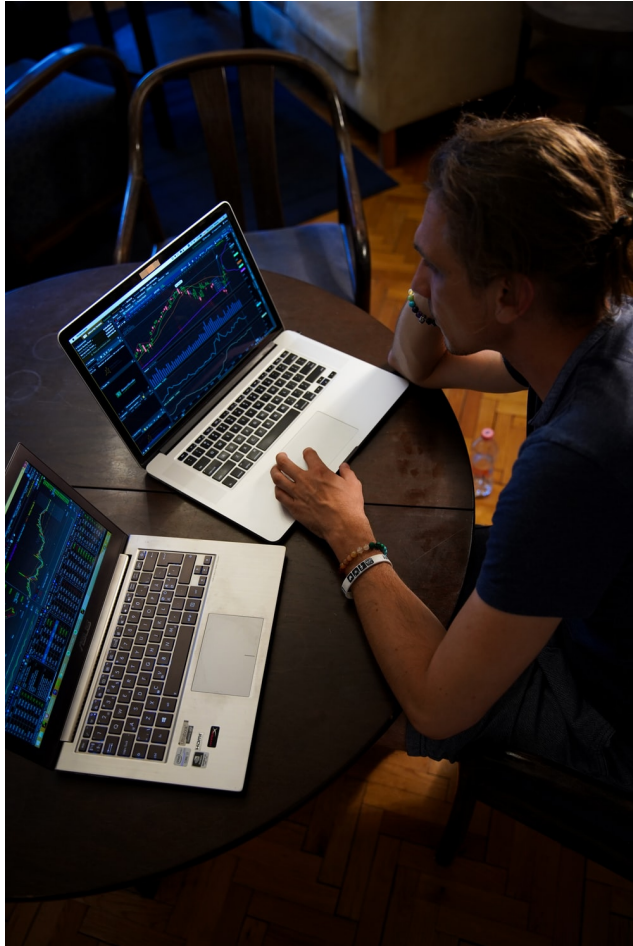
→ **Business-Anforderungen** mit **möglichen Daten** korrelieren

→ **Realistische** Anforderungen stellen

→ **Interdisziplinäre** Zusammenarbeit besonders wichtig



CRISP-DM → Cross Industry Standard Process for Data Mining



2) Data Understanding (Datenverständnis)

- Data Profiling = grundlegende Datenanalyse und Exploration
 - Bestimmung von Häufigkeiten, Wertebereichen, Korrelationen und Verteilungen
 - Analyse der Datenqualität (z.B. fehlende Werte, Ausreißer, Aktualität)
- Welche Arten von Datenquellen kommen in Frage?
 - Lexika (z.B. Wikipedia)
 - Redaktioneller Content (kuratiert)
 - User Generated Content

Vorab Data Engineering (Daten stehen bereits zur Analyse gut einlesbar zur Verfügung)

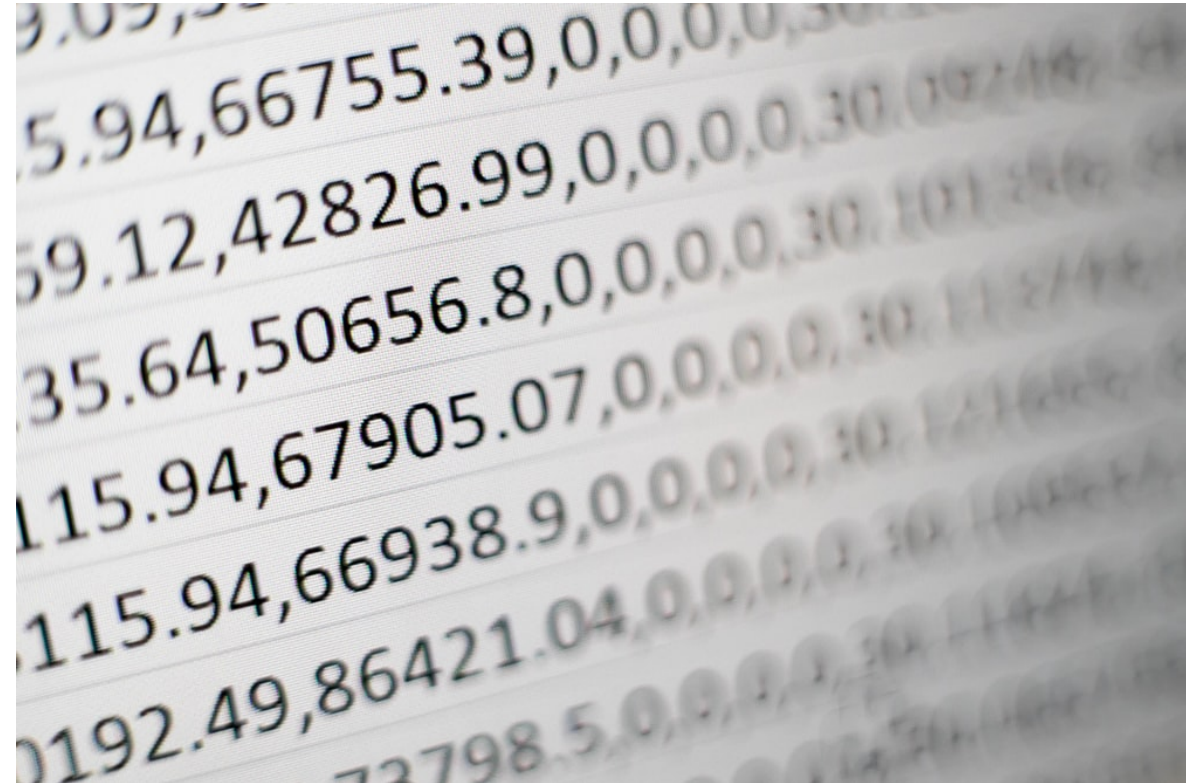
CRISP-DM → Cross Industry Standard Process for Data Mining

3) Data Preparation (Datenvorbereitung)

- Vorbereitung bzw. Aufbereitung der Input-Daten für das Data Mining
- Auswahl von Tabellen und Attributen, Festlegung von Filterbedingungen
- Transformation von Wertebereichen (z.B. Diskretisierung, Normalisierung)
- Datenbereinigung (z.B. Nullwerte, Ausreißer)

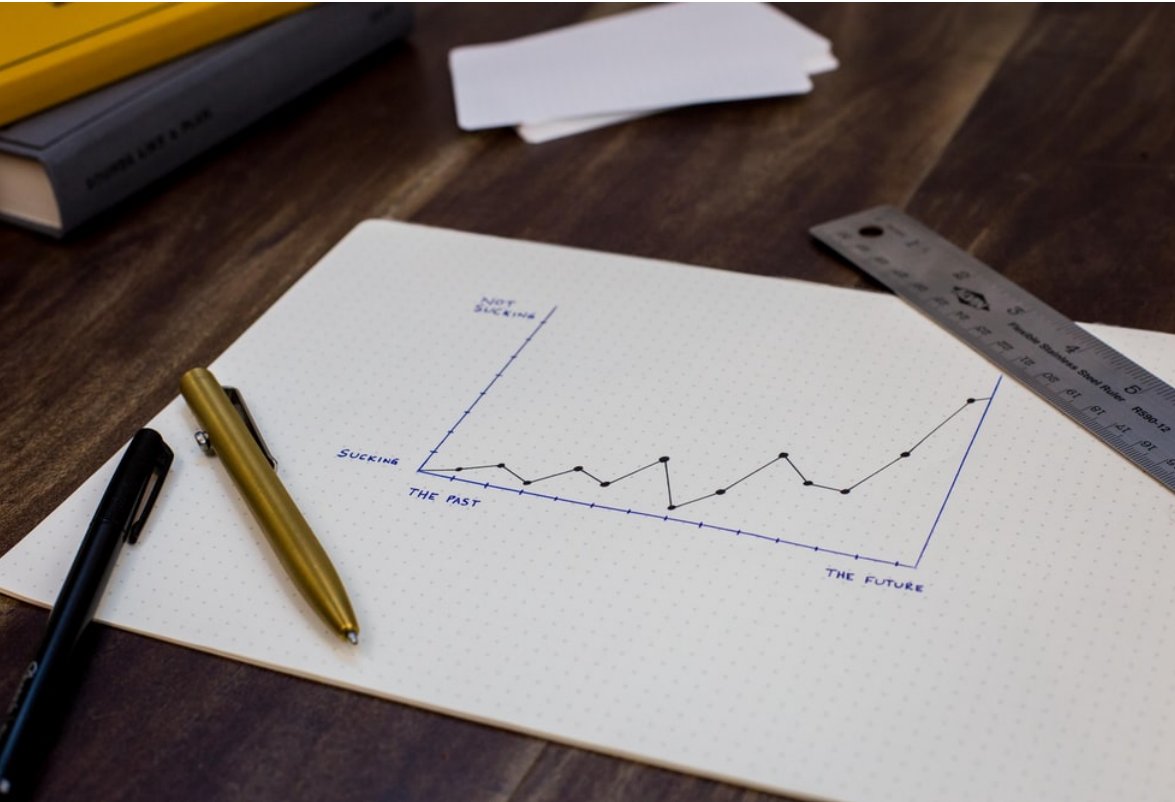
→ Analysemethoden

- (Deskriptive) Statistik
- Clustering, Selbstordnung
- Regression, Klassifikation und Anreicherung
- Zeitreihen und Trends
- Vorhersagen



<https://datasolut.com/crisp-dm-standard/#data-preparation>

CRISP-DM → Cross Industry Standard Process for Data Mining



4) Modeling (*eigentliches* Data Mining)

- Unterschiedliche Modellierungstechniken
- Nutzung von Data Mining zur Erstellung eines Modells (ggfs. erneut Data Preparation erforderlich)

→ Jonas Rende & Thomas Stadelmann: Machine Learning-Model (Sentiment)

→ Christian Winkler: Trend-Detection-Modell

CRISP-DM → Cross Industry Standard Process for Data Mining

5) Evaluation (Bewertung)

- Modell ist erstellt und muss nun auf seine Qualität hin bewertet werden
- Berechnung von Qualitätsmetriken
- Überprüfung der Rahmenbedingungen und Einsatzfähigkeit
- Erfüllt das Modell den Geschäftszweck?

→ Relevant für Evaluation: Verständliche und ansprechende **Aufbereitung der Analyseergebnisse**

→ **Data Storytelling & Visualisierung**



CRISP-DM → Cross Industry Standard Process for Data Mining



- 6) **Deployment** (Anwendung, Operationalisierung)
- Anwendung des Modells im Entscheidungsprozess
 - Integration von Scorings oder Regeln in operative Prozesse zur manuellen oder automatisierten Entscheidungsfindung
 - Kontinuierliche Darstellung und Einbettung in operative Entscheidungsfindungsprozesse, z.B. mittels **Dashboards**

Business Understanding: Use Case

02

Use Case "Automobilhersteller" → Branchenreports



Themen Branchen Dienstleistungen E

Automobilindustrie

Als vertrauensvoller Partner unterstützen wir die Automobilindustrie bei der proaktiven Gestaltung der Mobilität von morgen.

Vom Auto zum Algorithmus

Autos zu bauen reicht nicht. Entscheidend ist, Bedürfnisse zu verstehen und zu bedienen.



Branchenreport Ausgabe 2020

AUTOMOBIL | GENIOS BranchenWissen Nr. 11 vom 10.11.2020

Genios 2021; KPMG 2021; VDA 2021; Deutsche Mittelstandsfinanz GmbH 2021



Deutsche
Mittelstandsfinanz

Branchenreport Automotive

Die deutsche Automobilindustrie
und das Ende des „Weiter so“!



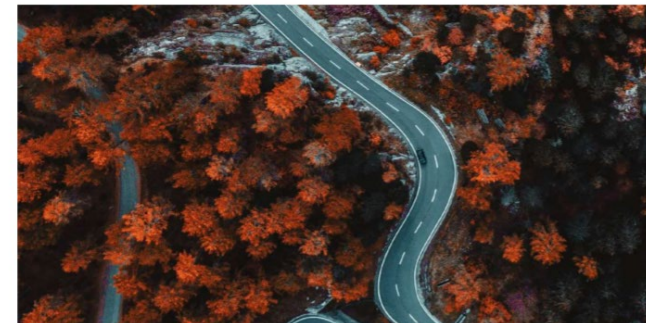
Forschungsvereinigung
Automobiltechnik



Verband der
Automobilindustrie

Roadmap für die Automobilität der Zukunft

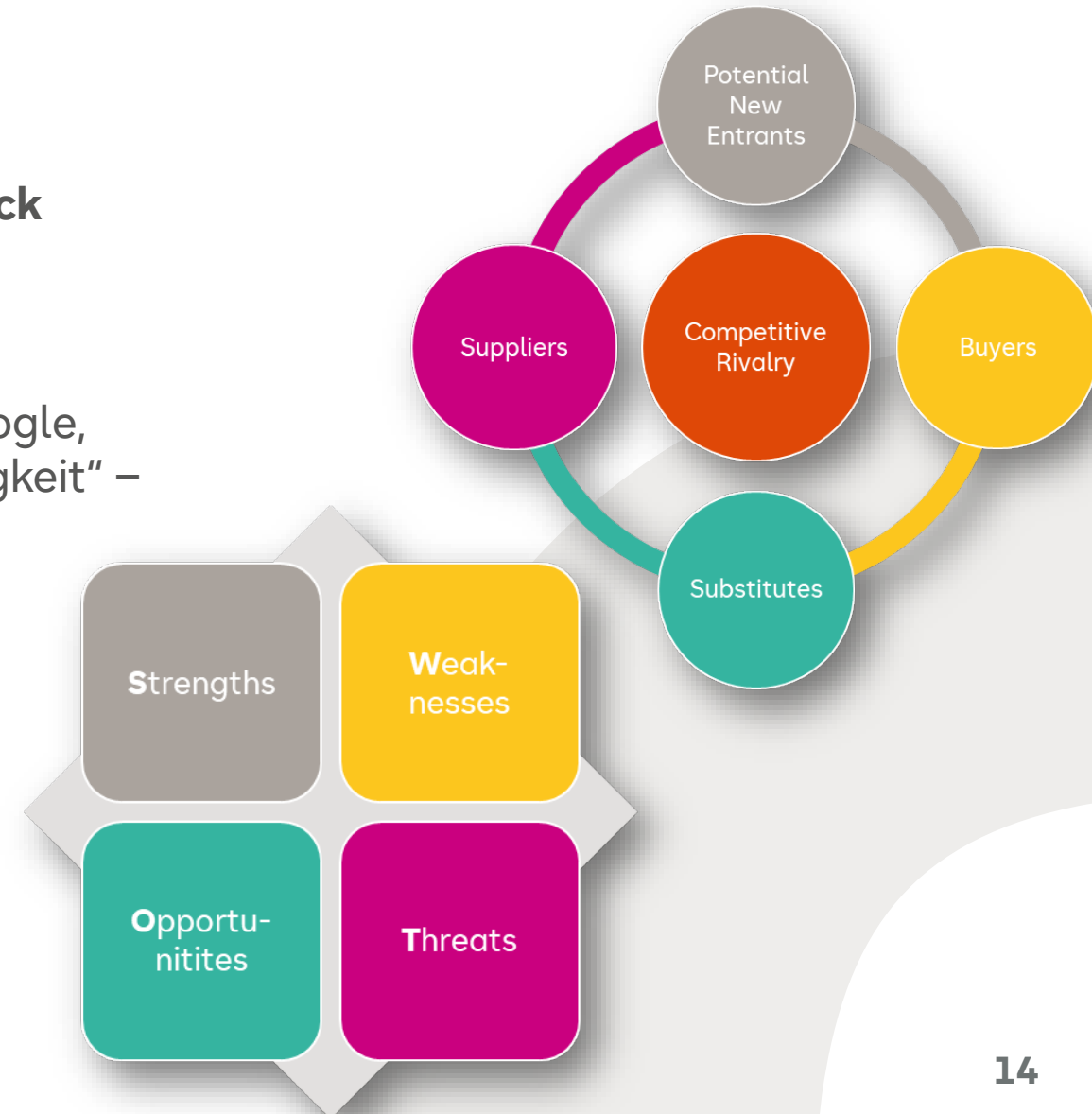
Grundstein einer Innovationspartnerschaft



Ergebnisse aus Beratungsanalysen

Ausgangssituation

- Ergebnisse aus Branchenreports zeigen: **Innovationsdruck** steigt für gesamte Branche
- Analyse durch Unternehmensberatung zeigt:
 - massiver **externer Druck** (Porter's Five Forces) → bestehender Wettbewerb, neue Wettbewerber (Google, Apple), Kulturwandel bei Verbrauchern („Nachhaltigkeit“ – Car Sharing, E-Mobility)
 - **Interne Schwächen** (SWOT, insb. fehlendes Verständnis über „Customer's Voice“)
- Wunsch aus Management nach
 - „Schnellen Erkenntnissen“ (keine Zeit für Wellen-/Panelbefragungen)
 - Identifikation von Trendthemen
 - Ableitung konkreter Implikationen für Positionierung
 - „Data driven Insights mittels AI“



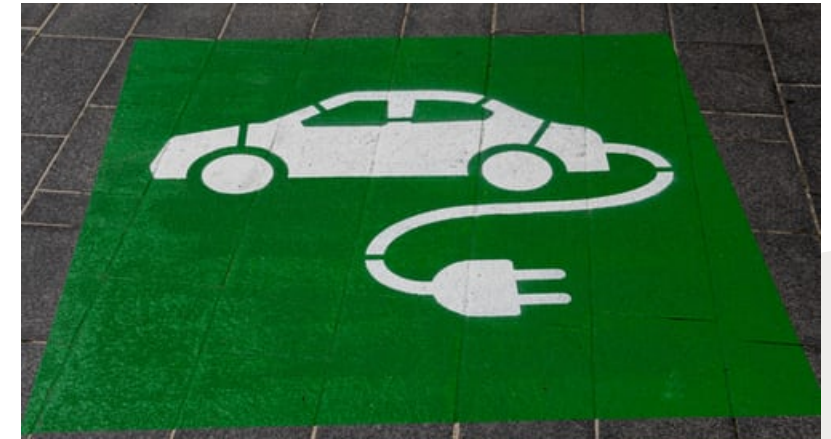
Business Understanding – Use Case

→ **Commitment** bzgl. konkreter Fragestellungen der Stakeholder:

- Welche Themen werden von (potentiellen) Kunden bzw. Endverbrauchern bzgl. **innovativer Technologien** im Automobil-Sektor diskutiert?
 - Können **Trends** im Bereich moderner Technologien identifiziert werden?
 - Wie stehen Verbraucher bestimmten Themen gegenüber (**Assoziationen, Einstellungen**)?
 - Differenzierungspotential über **Mobility als Service**?
- Ziel: Besseres Verständnis über relevante Themen bei sog. „**Early Adopters**“, d.h. **Technologie-affinen** Verbrauchern für Positionierung und externe Kommunikation

→ Wichtig: **Erwartungsmanagement** bei internen Stakeholdern

- Unrealistische Erwartungen frühzeitig „einfangen“
- Iterativer, interdisziplinärer Analyseprozess



Soziale Netzwerke

03

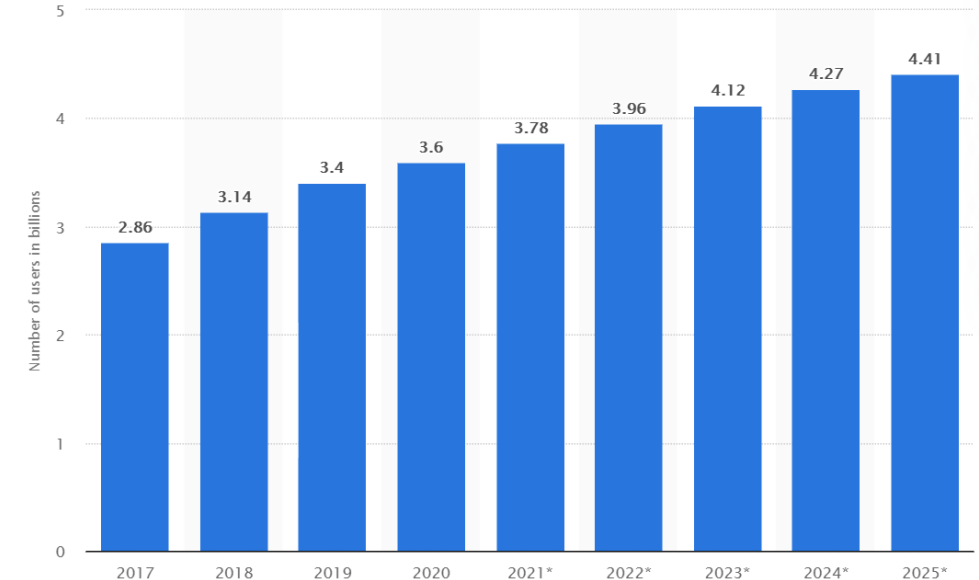
Unterschiedliche soziale Netzwerke

➔ Vielzahl an Social Media Kanälen

- Facebook (im Downturn)
- Instagram (nur Bilder, Star-Kult)
- Tiktok (nur Videos)
- Youtube (nur Videos)
- Reddit (Rising Star?)
- Snapchat (Instant Messaging)
- WhatsApp (Instant Messaging)

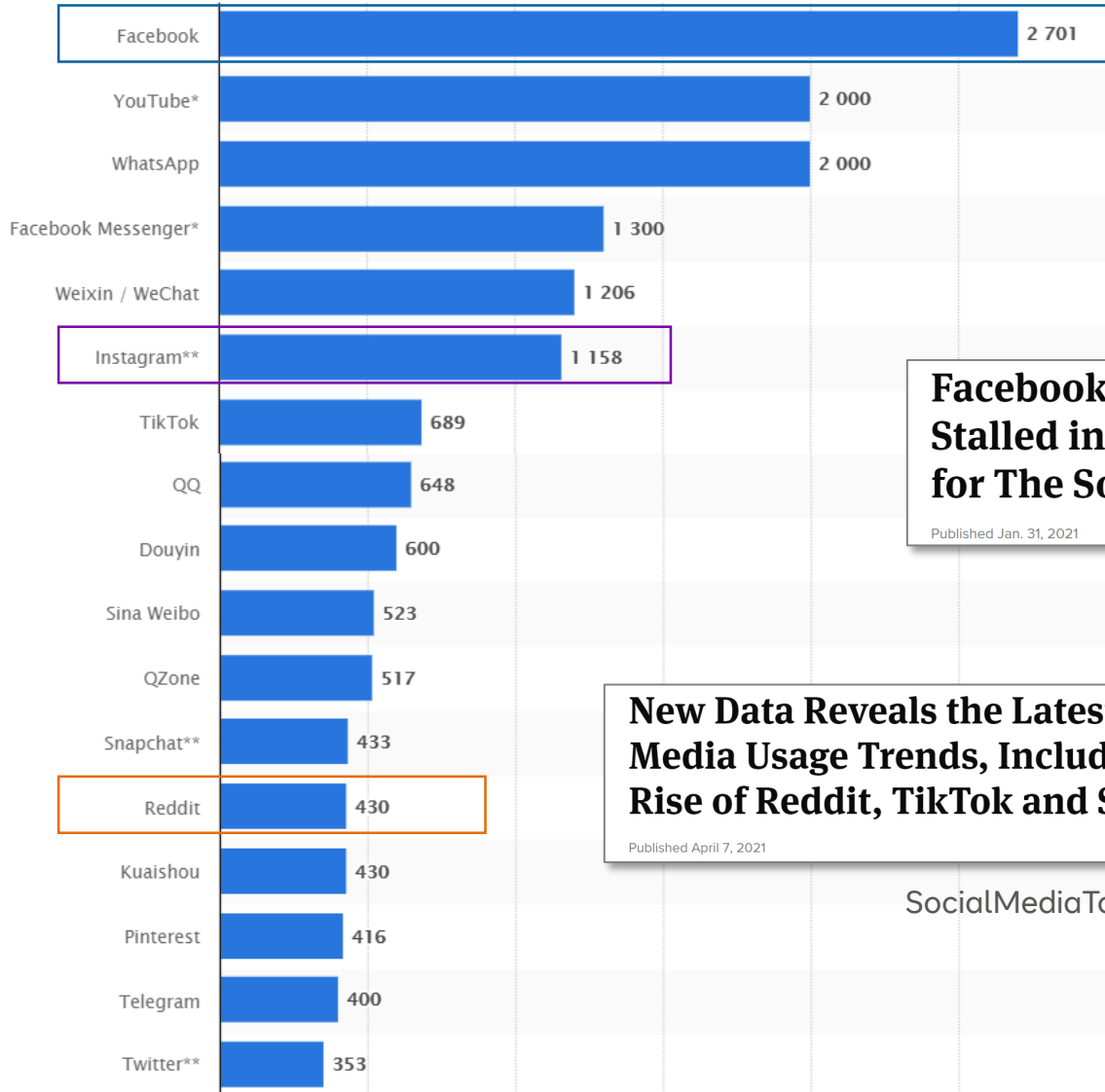


Number of social network users worldwide from 2017 to 2025
(in billions)



Statista (2021)

Soziale Netzwerke – aktuelle Entwicklungen



Wie sehr die Facebook-Dominanz quer durch alle Altersgruppen seit 2015 gesunken ist

28. April 2021, Autor: Michael Kroker

Seit 2015 ist Facebook als meistgenutztes soziales Netzwerk von gut zwei Drittel auf erstmals weniger als Hälfte aller Social-User gesunken.

Wirtschaftswoche (2021)

Facebook's Daily Active Usage Has Stalled in the US - A Sign of Concern for The Social Network?

Published Jan. 31, 2021

SocialMediaToday (2021)

New Data Reveals the Latest Social Media Usage Trends, Including the Rise of Reddit, TikTok and Snapchat

Published April 7, 2021

SocialMediaToday (2021)

HOME > TECH

Facebook reported a decline of 2 million daily active users in the US and Canada

BusinessInsider (2020)

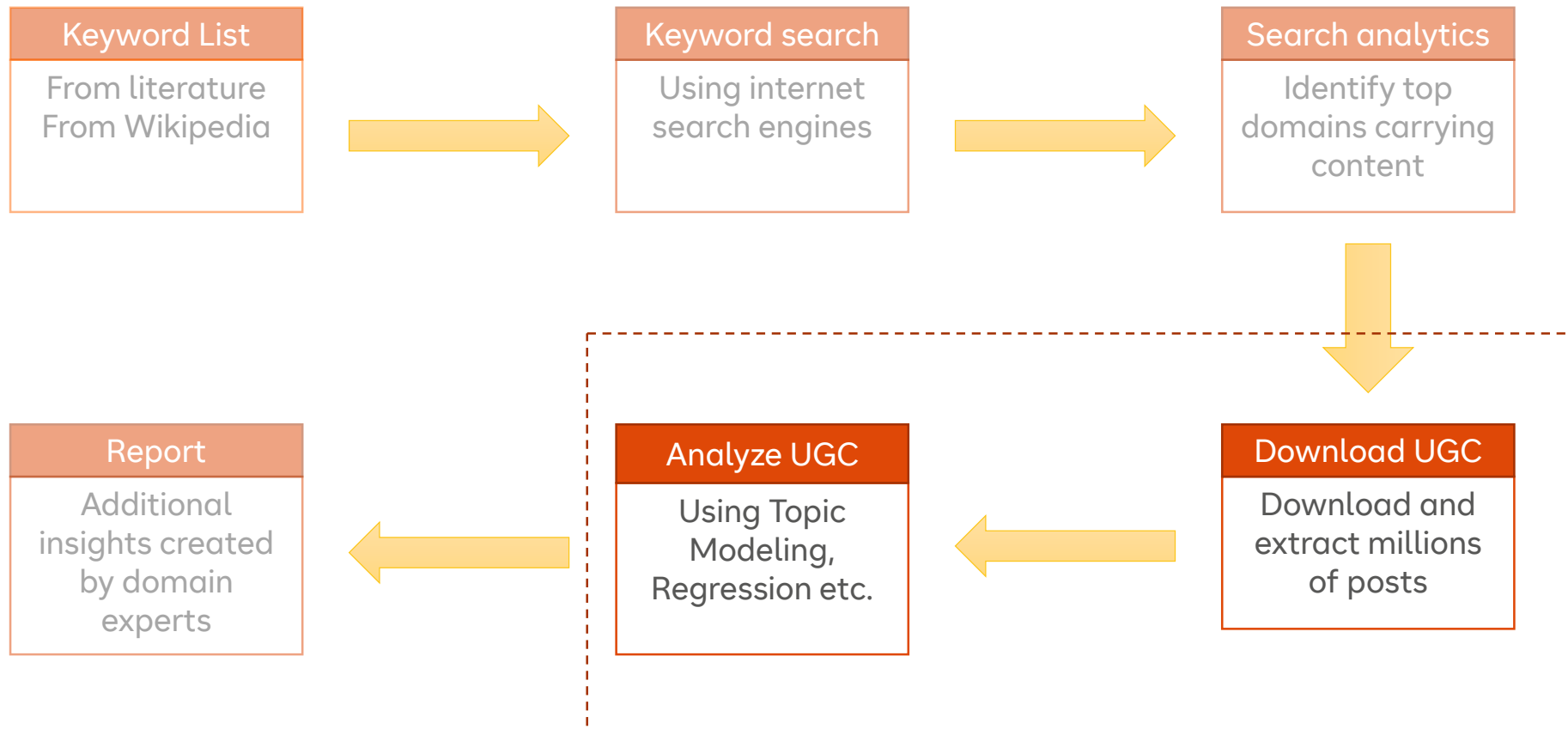
- ➔ Facebook im Downturn, verliert User
- ➔ Instagram wächst seit 2013 (zunehmend langsamer)
- ➔ **Reddit** unter den Top 10-Websites (7) in den USA (Instagram 18) vs. Top 20 (18) global (Instagram 22) gemessen an MAUs und DAUs

Alexa (2021), Hypestat (2021), pewresearch.org (2019)

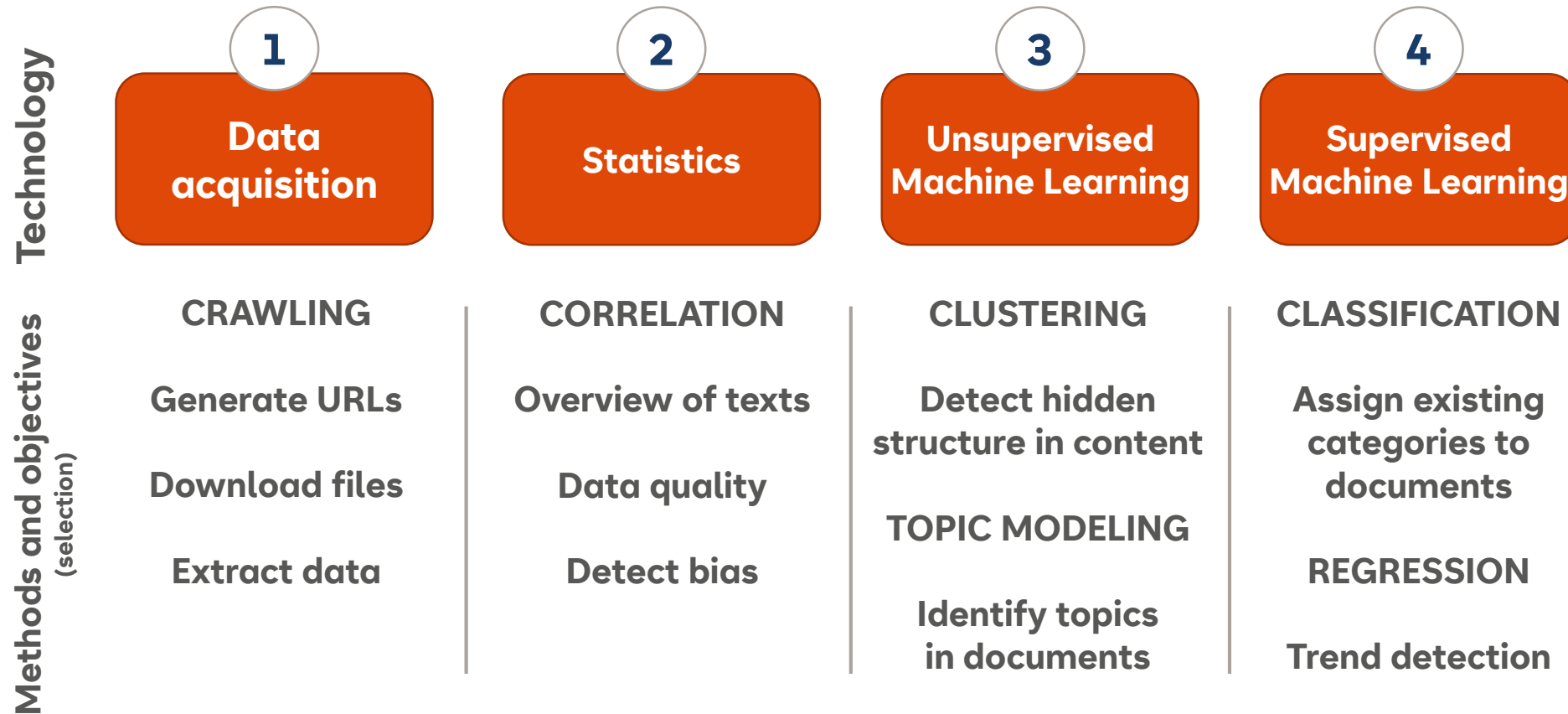
Projektablauf einer UGC-Analyse

04

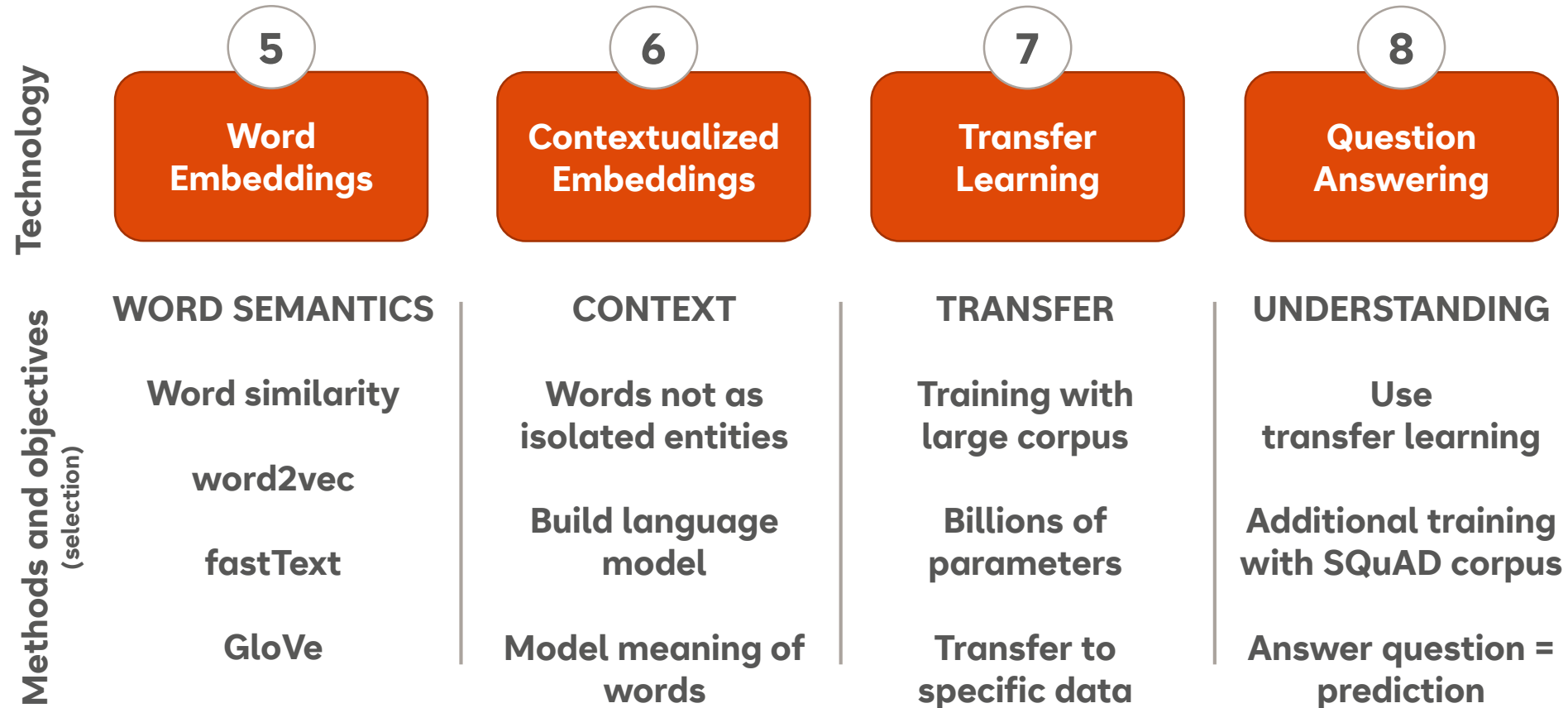
Projektphasen



Technologien – Teil 1



Technologien – Teil 2



Unser Beispiel: Subreddit „Technology“

05

Beispiel "r/Technology"

Inhaltlich

- Inhaltlich relevanter Content
- Aktualität
- Ergiebige Auswertungsdimensionen

Technisch


- API vorhanden
- Download erlaubt

Über diese Community

Subreddit dedicated to the news and discussions about the creation and use of technology and its surrounding issues.

10.7m
Mitglieder

14.7k
Online

 Am 25. Jan. 2008 erstellt

Nach Flair filtern

TechSupport

Social Media

Hardware

Business

Society

Networking/Telecom

Transportation

Politics

Space

Biotechnology

Security

Energy

Privacy

Robotics/Automation

Nanotech/Materials

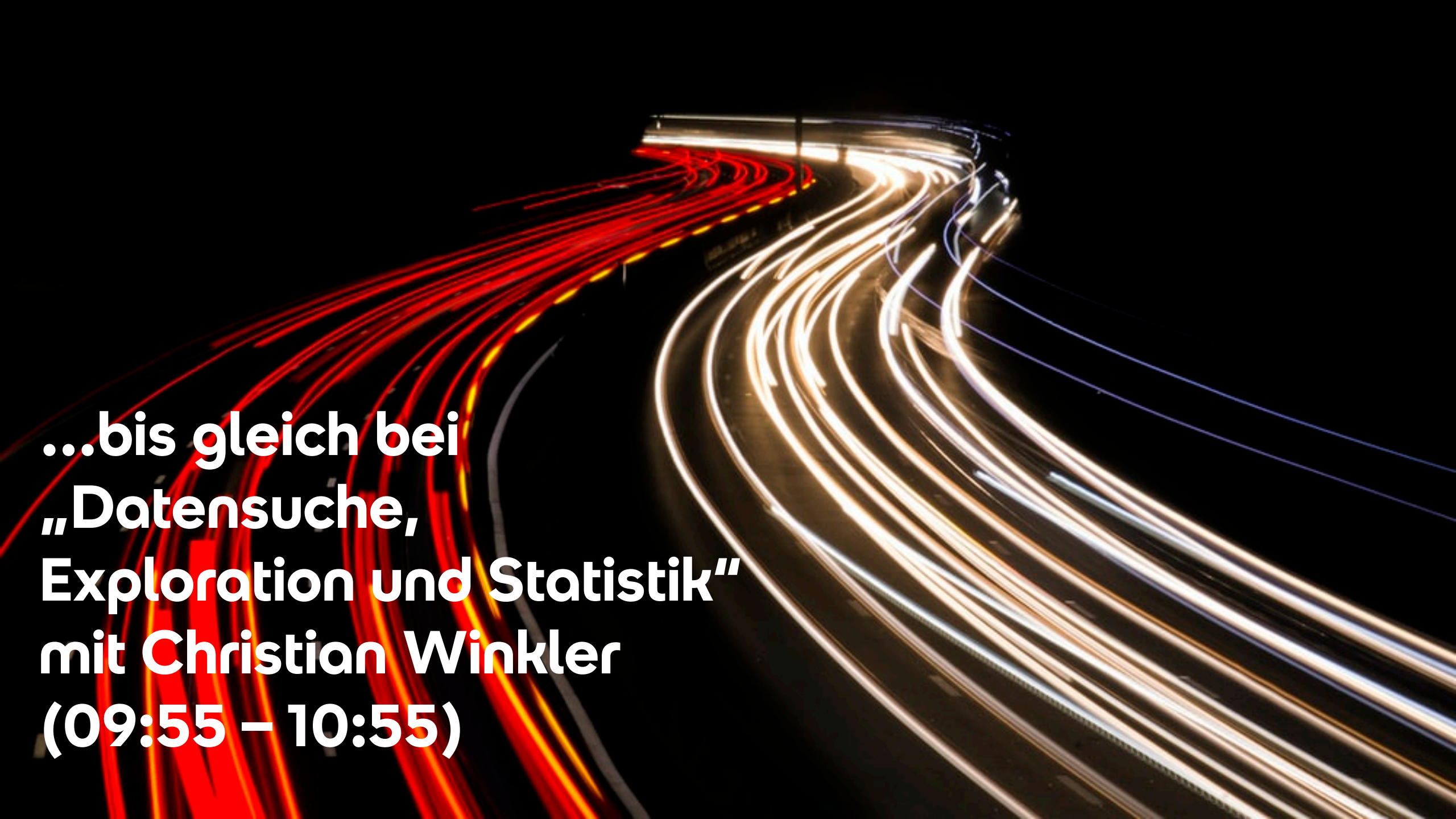
Crypto

Artificial Intelligence

Software

Beispiel: Hierarchische Struktur bei Reddit





**...bis gleich bei
„Datensuche,
Exploration und Statistik“
mit Christian Winkler
(09:55 – 10:55)**