

Das Beratungsangebot des Kompetenzzentrums für Lizenzierung (KfL) zu Text-and-Data-Mining („TDM-Support“)

Folien: **Dr. Tillmann Dönicke**

Vortrag: **Niklas Spies**

Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen)

Kompetenzzentrum für Lizenzierung (KfL)

Ein DFG-gefördertes Kooperationsprojekt

www.fid-lizenzen.de

Coffee Lecture der AG-FID

13.12.2024

Was versteht man unter TDM?

Hintergrund...

Verfügbarkeit großer digitaler Datenmengen und computerbasierter Analysemethoden nimmt zu -> neue Wege für die Wissenschaft!

Data-Mining ...

- soll möglichst autonom und effizient Muster innerhalb großer Datensätze identifizieren
- nutzt Methoden aus der Statistik, dem maschinellen Lernen und der künstlichen Intelligenz

Text-Mining ...

- richtet sich speziell an die Verarbeitung großer Textmengen

Bedarfsanalyse

Umfrage 2022

> 80% der FIDs Bedarf an Unterstützungsdienstleistungen im Bereich TDM
-> AP 9 „TDM-Support“ im aktuellen KfL-Förderzeitraum (01.04.2023-31.03.2025), das von der SUB in Abstimmung mit SBB und BSB bearbeitet wird

Umfrage 2024

Anfang 2024 wurde eine weitere Umfrage unter den FIDs durchgeführt:

- 18 der 42 FIDs haben teilgenommen
- etwa gleichverteilt: TDM für generelle Datenaufbereitung, TDM für Projektanträge, TDM ohne bestimmten Anlass, kein Interesse an TDM
- überwiegende Mehrheit aus Geistes- und Kulturwissenschaften
-> großes Interesse an Textverarbeitung, wenig Erfahrung mit TDM
- gewünschte Beratungsfelder: Aufbereitung von Daten zu Analysezielen und statistische Analysen
- gewünschte Kompetenzbereiche: Data Science und Lizenzrecht

Bereiche des TDM-Supports

Von unstrukturierten Rohdaten zu nachnutzbaren Forschungsergebnissen

1. Problemverständnis

- Definition der Projektziele und -fragen
- Erstellung eines Projektplans

2. Datenverständnis

- Datensammlung und -exploration
- Beschreibung der Daten
- Bewertung der Datenqualität

3. Datenvorbereitung

- Datenbereinigung
- heterogene Formatierung
- Merkmalsextraktion
- Datenauswahl
- Annotation/Tagging

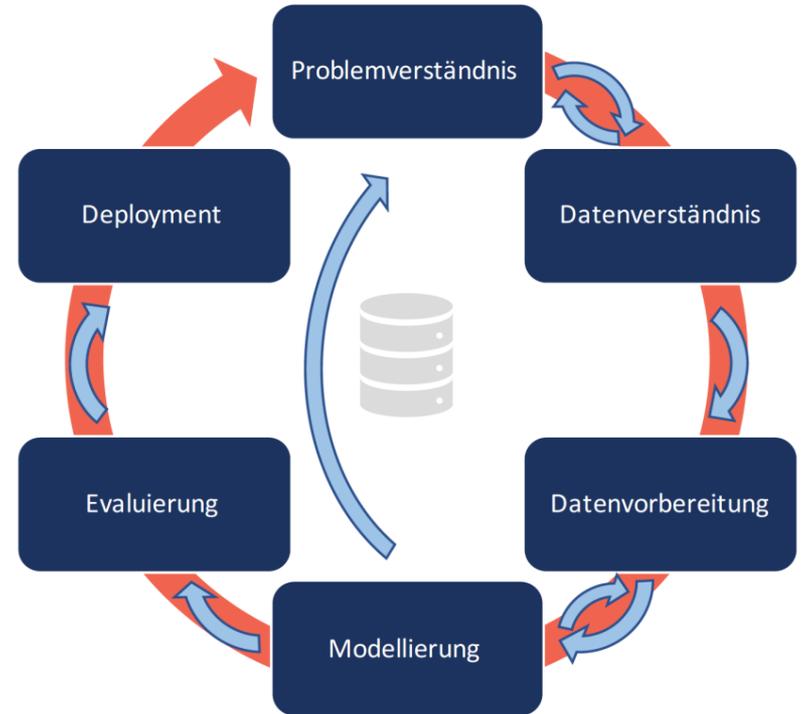


Abb. 1: Life cycle eines TDM-Projekts nach dem CRISP-DM Modell

Bereiche des TDM-Supports II

Von unstrukturierten Rohdaten zu nachnutzbaren Forschungsergebnissen

4. Modellierung

- Auswahl von Modellierungstechniken
- Aufbau und Training von Modellen
- Modellbewertung und -optimierung
- statistische Datenanalyse

5. Evaluierung

- Evaluierung der Modelle hinsichtlich Effektivität und Genauigkeit

6. Deployment

- Bereitstellung von Modellen als operative Anwendung
- Bereitstellung aufbereiteter Daten
- Visualisierung von Ergebnissen

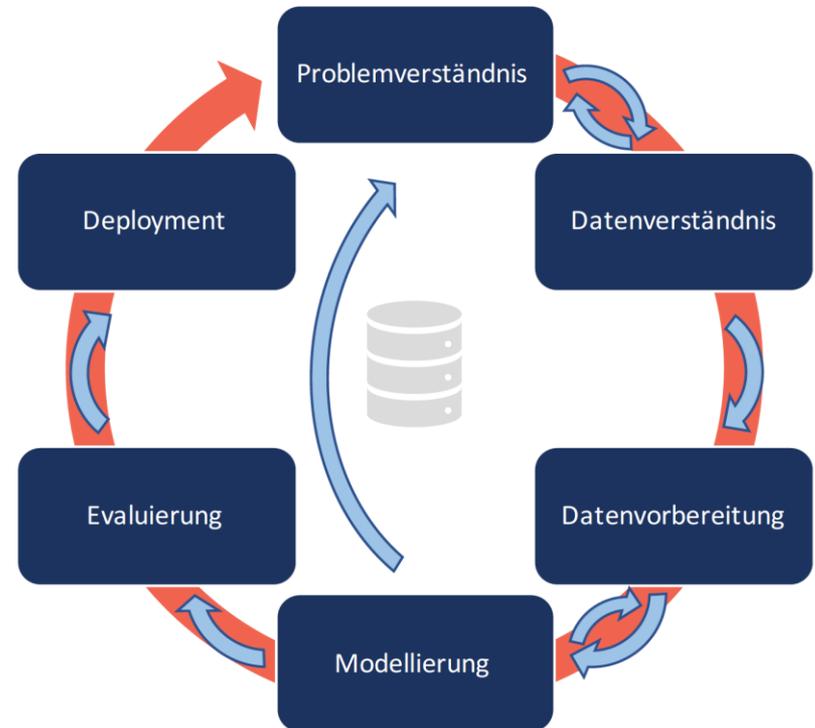


Abb. 1: Life cycle eines TDM-Projekts nach dem CRISP-DM Modell

Angebotene Leistungen

Fachliche Beratung der FIDs bei ...

- Planung und Durchführung von TDM-Projekten
- Auswahl geeigneter Daten und Werkzeuge
- Interpretation von Ergebnissen

Technische Unterstützung der FIDs bei ...

- Einrichtung und Nutzung von TDM-Ressourcen
- Behebung technischer Probleme
- ABER: keine Programmierarbeiten

Weiterführende Unterstützung der FIDs bei ...

- Entwicklung und Optimierung von TDM-Methoden
- Verbesserung der Methodeneffizienz und -skalierbarkeit

Rechtliche Beratung der FIDs bei ...

- Fragen zum Urheber- und Datenschutzrecht

Ablauf

1. Informationserhebung

Erstbearbeiter*in befragt Anfragersteller*in zu Daten, Software und TDM-Vorhaben.

2. Dateninspektion

TDM-Expert*in macht sich mit Daten / Datenstichprobe von Anfragersteller*in vertraut.

3. Beratung

TDM-Expert*in versucht, eigentliches Problem von Anfragersteller*in zu klären.

4. Feedback

Anfragersteller*in wird gebeten, an Nachbefragung durch Service-Mitarbeiter*in teilzunehmen.

Ablauf II

Die Kommunikation läuft via Ticket-System.

Service-Mitarbeiter*innen:

- 3 TDM-Expert*innen in Data Science, insbes. Textverarbeitung/Textanalyse
- 1 TDM-Expert*in in Lizenzrecht

Derzeitige und künftige Entwicklungen

Aufnahme des Betriebs: 15.07.2024

Bisherige Anfragen: 0

Eine fortlaufende Evaluation (z. B. durch kurze Nachbefragungen) und ggf. Anpassungen sind geplant.

Informationen und Kontakt

FID-Infos / Beratung Text-and-Data-Mining („TDM-Support“)

https://www.fid-lizenzen.de/TDMsupportInfo_1.0_final.pdf

(oder als Unterpunkt auf <https://www.fid-lizenzen.de/fid-infos>)

E-Mail-Adresse des TDM-Supports: tdm-support@fid-lizenzen.de

Hilfreiche Informationen für die Beratung:

- Forschungsdesign
- Format, Umfang und Qualität der Daten
- Software
- TDM-Vorhaben
- Beispiel-/Dummy-Daten