

Duale Hochschule Baden-Württemberg Mannheim

## **Dokumentation**

# **Entwicklung eines NLP-Tools zur Textzusammenfassung und -klassifizierung**

## **Studiengang Wirtschaftsinformatik**

**Studienrichtung Data Science**

Dozenten:	Michael Lange, Enzo Hilzinger
Verfasser:	Bonn Lukas, Amina Uicker-Darwish, Jan Rüdts, Aymane Bougern
Matrikelnummer:	5856761, 8278962, 1737304, 1552312
Kurs:	WWI20DSA
Studiengangsleiter:	Prof. Dr.-Ing. habil. Dennis Pfisterer
Bearbeitungszeitraum:	08.05.2023 – 24.05.2023

# 1 Projektstrukturplan

Zur weiteren Planung des Projekts wird im Folgenden der in Abbildung 1.1 dargestellte Projektstrukturplan vorgestellt.

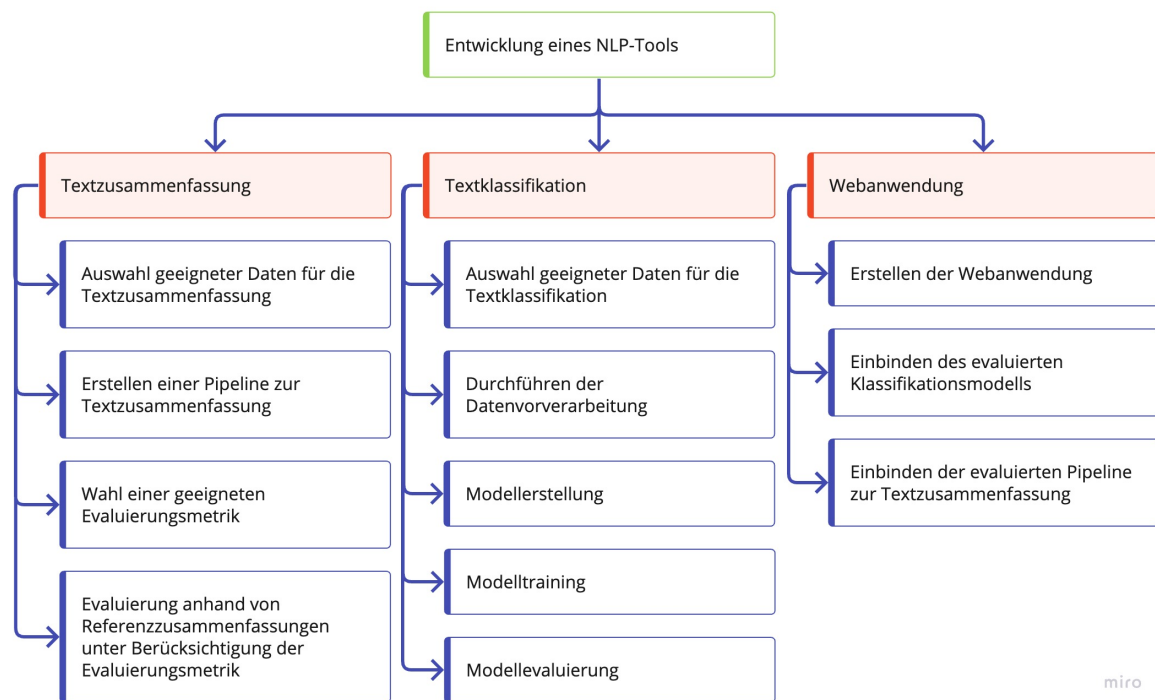


Abbildung 1.1: Eigene Darstellung: Grober Projektstrukturplan

Durch diesen können die Aufgaben, welche im Rahmen des Projekts anfallen, auf eine übersichtliche Weise abgebildet werden, sodass aufbauend darauf die weiteren Projektschritte umgesetzt werden können (vgl. Felkai et al., 2013, S. 217). Die hierbei definierten Teilaufgaben werden *Arbeitspakete* genannt (vgl. Peters et al., 2021, S. 68). Innerhalb des betrachteten Projekts wird zur detaillierten Gliederung des Projektstrukturplans ein objektorientierter Zerlegungsansatz gewählt. Bei diesem findet die Zerlegung des Projekts in Objekte bzw. Teilobjekte unter Berücksichtigung der technischen Struktur des Gesamtsystems statt (vgl. Felkai et al., 2013, S. 219).

Da innerhalb des zu entwickelnden NLP-Tools, wie bereits im Pflichtenheft dargestellt, verschiedene Bestandteile (*Objekte*) von technischer Seite aus von Relevanz sind, wird ein solches Vorgehen als sinnvoll erachtet.

Folgende Objekte werden mit den dazugehörigen Arbeitspaketen zur Gliederung des Projektstrukturplans definiert:

- **Textzusammenfassung**

Als erste Teilaufgabe der Textzusammenfassung ist die Auswahl eines geeigneten Datensatzes von Relevanz, um dadurch die Durchführung einzelner weiterer Aufgaben zu ermöglichen. Zudem ist es wichtig, eine Pipeline zu erstellen, welche die Zusammenfassung eines gegebenen Texts ermöglicht. Detailliertere Informationen zu den einzelnen Schritten innerhalb dieser Pipeline sind im Pflichtenheft dargelegt und werden im detaillierten Projektstrukturplan herausgearbeitet. Zudem ist als weiteres Arbeitspaket die Wahl einer Metrik essenziell, um mit dieser die Textzusammenfassung zu evaluieren. Erst dadurch kann die Qualität dieses Bestandteils bewertet werden. Die konkrete Evaluierung anhand von Referenzzusammenfassungen ist als weiteres Arbeitspaket definiert.

- **Textklassifikation**

Im Rahmen der Textklassifikation wird, wie auch bei der Zusammenfassung, die Auswahl geeigneter Daten als wesentliches Arbeitspaket angesehen, um aufbauend darauf Vorverarbeitungsschritte auf diesen Daten als Teilaufgabe durchführen zu können. Als Kern der Klassifikation werden zudem die Modellerstellung, das Modelltraining und die Modellevaluierung als essentieller Bestandteil des Projektstrukturplans angesehen, um dadurch die Anforderung an die Zurverfügungstellung einer Klassifikationsfunktion erfüllen zu können.

- **Webanwendung**

Das dritte Objekt bildet die Webanwendung, um dadurch eine Schnittstelle für den Benutzer zu bieten, sodass dieser durch die erarbeiteten Themen einen Mehrwert ziehen kann. Im Zuge dessen wird die Erstellung der Webanwendung als ein Arbeitspaket angesehen. Darüber hinaus werden das Einbinden des Klassifikationsmodells, welches im Rahmen der Evaluierung als bestes Modell identifiziert wurde, wie auch das Einbinden der Zusammenfassungsfunktion in die Webanwendung jeweils als Teilaufgabe betrachtet. Erst durch diesen Prozess kann die Anforderung des Auftraggebers, eine Schnittstelle bereitzustellen, mithilfe derer der Benutzer auf die NLP-Funktionalitäten zugreifen kann (*siehe Lastenheft*), erfüllt werden.

# Literaturverzeichnis

- Felkai, R., Beiderwieden, A., Felkai, R., & Beiderwieden, A. (2013). Planen des gesamten Projekts. *Projektmanagement für technische Projekte: Ein prozessorientierter Leitfaden für die Praxis*, 215–267.
- Peters, T., Schelter, N., Peters, T., & Schelter, N. (2021). Antworten auf die Kontrollfragen. *Kompakte Einführung in das Projektmanagement: Mit vielen praxisnahen Beispielen und modernen didaktischen Instrumenten*, 181–187.