













Datensatzverarbeitung (engl. dataset pipeline)														
Datensatzakquisition (engl. dataset acquisition)					Datensatzsichtung (engl. dataset inspection)					Datensatzaufbereitung (engl. dataset preparation)				
														
Datensatzrechenrche (engl. dataset research)	Datensatzsammlung (engl. dataset collection)		Datensatzprüfung (engl. dataset check)	Datensatzauswahl (engl. dataset selection)	Datenstrukturanalyse (engl. data structure analysis)	Explorative Datenanalyse (engl. exploratory data analysis) In der EDA werden Textdaten untersucht, um Muster, Qualitätsprobleme und Strukturen zu erkennen. <small>(perplexity.ai)</small>					Datensatzbereinigung (engl. dataset cleaning)		Datensatzvalidierung (engl. dataset validation)	
Suchstrategie (engl. search strategy)	Download (engl. data download)	Speicherung (engl. data storage)				Analyse der strukturierten Daten (engl. analysis of structured data)			Analyse der unstrukturierten Daten (engl. analysis of unstructured data)		Fehlwertbehandlung (engl. missing value handling)	Duplikatentfernung (engl. duplicate removal)		
						Ortsdaten (engl. location data)	Zeitdaten (engl. time data)	Bewertungen (engl. rating data)	Fehlwerterkennung (engl. missing value detection)		Duplikaterkennung (engl. duplicate detection)	Textlängenanalyse (engl. text length analysis)		
														
Rohzeichenfolgen (engl. raw strings)														
<u>Suchbegriffe</u> <ul style="list-style-type: none">„Beschwerden“„Kommentare“„NLP“„Kundenbeschwerden“„Online-Kommentare“„Produktbewertungen“		• lokale Spreicherung	• KI-Detektor https://github.com/Kishanjaisoorya/AI-Text-Detector-python			• Länder	• Zeitraumbestimmung		• NaNs (Not a Number)			• Imputation	Duplikate verzerren NLP-Modelle, daher müssen sie entfernt werden.	
<u>Anforderungen</u> <ul style="list-style-type: none">organischer Datensatzes						• Bundesstaaten	• Datumsanalyse		• NaTs (Not a Text)			• Listenweiser Fallausschluss „Strategies for handling missing data in text include imputation where missing values are estimated based on existing data and removal where incomplete observations are excluded from the analysis. These strategies help in ensuring the completeness and integrity of the text data [12].“ <small>(Upadhye, 2020, p. 207)</small>		