Projektrealisierung

Abschlusspräsentation – 27.07.23

Von Lukas Bonn, Amina Uicker-Darwish,
Jan Rüdt, Aymane Bouguern

Agenda

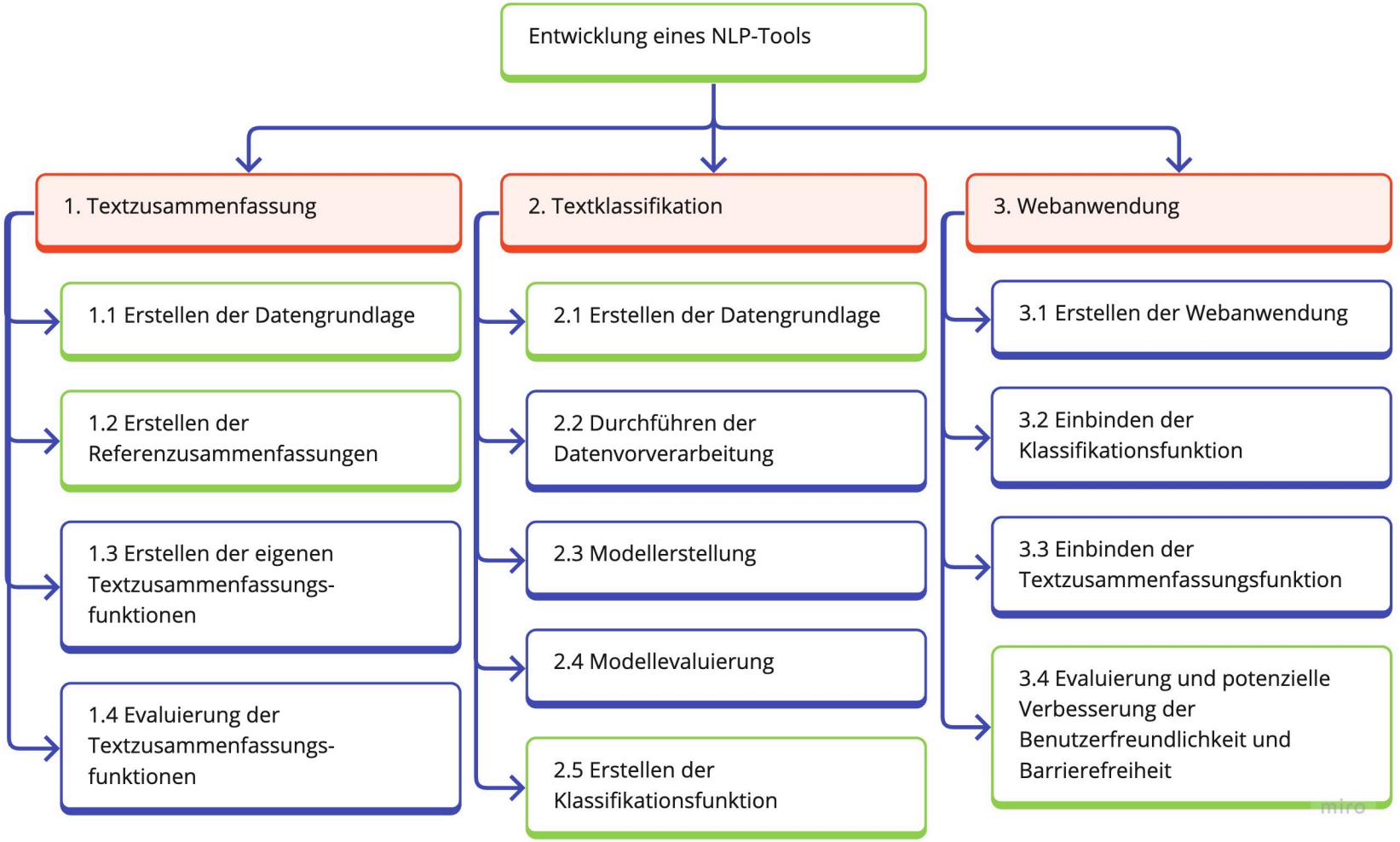
Recap

Implementierung + Erfüllung der Anforderungen

Ergebnis Demo

Abschluss: Zeitplan, Herausforderungen, Lessons Learned

Projektstrukturplan



5 Datensätze + Vorverarbeitung

Presidential Speeches

- Textsequenzen jeweils zwischen 500-800 Wörtern

CNN News

- Textlänge mind. 250 Wörter
- Entfernung von Überschriften wie z.B. „BAGHDAD, Iraq (CNN) “

Jurisdictions

- Extrahierung der Texte mit 500-800 Wörtern

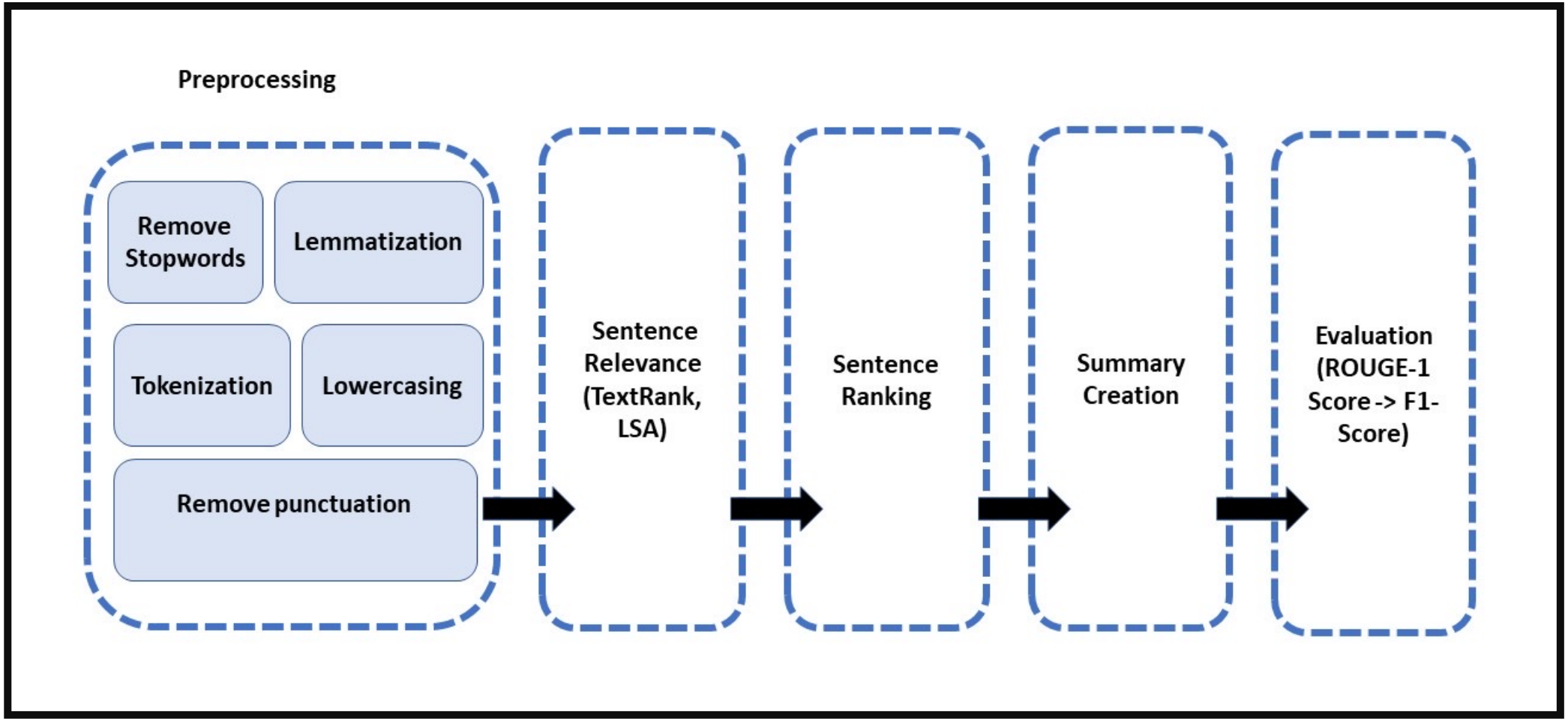
Literature

- Entfernung von Elementen wie Zeilenümbrüche, Mehrfachleerzeichen, Seitenzahlen, Autorinfos, etc.
- Texte auf 500-800 Wörter gekürzt

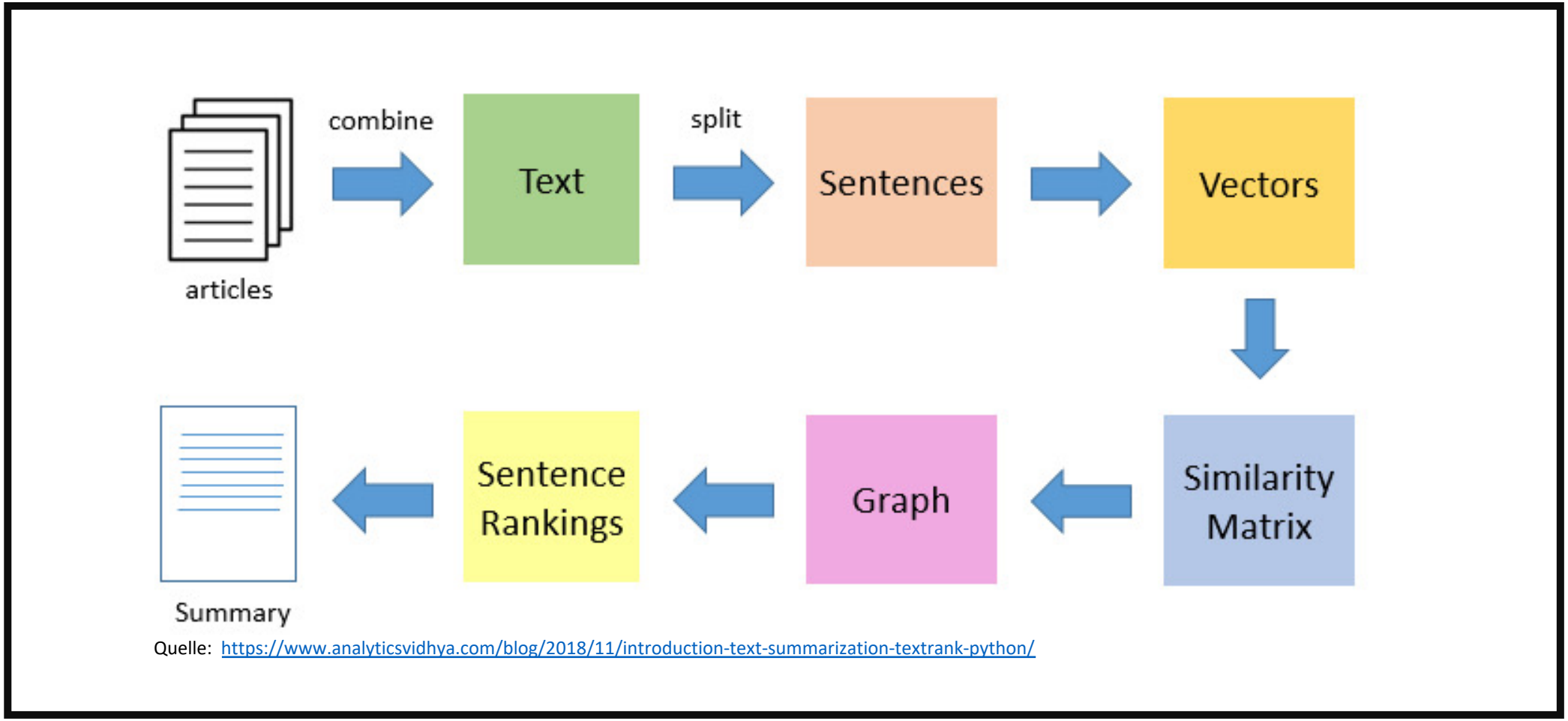
Blogs

- Extrahieurng von Sequenzen mit ca. 650 Wörtern

Implementierung Textzusammenfassung

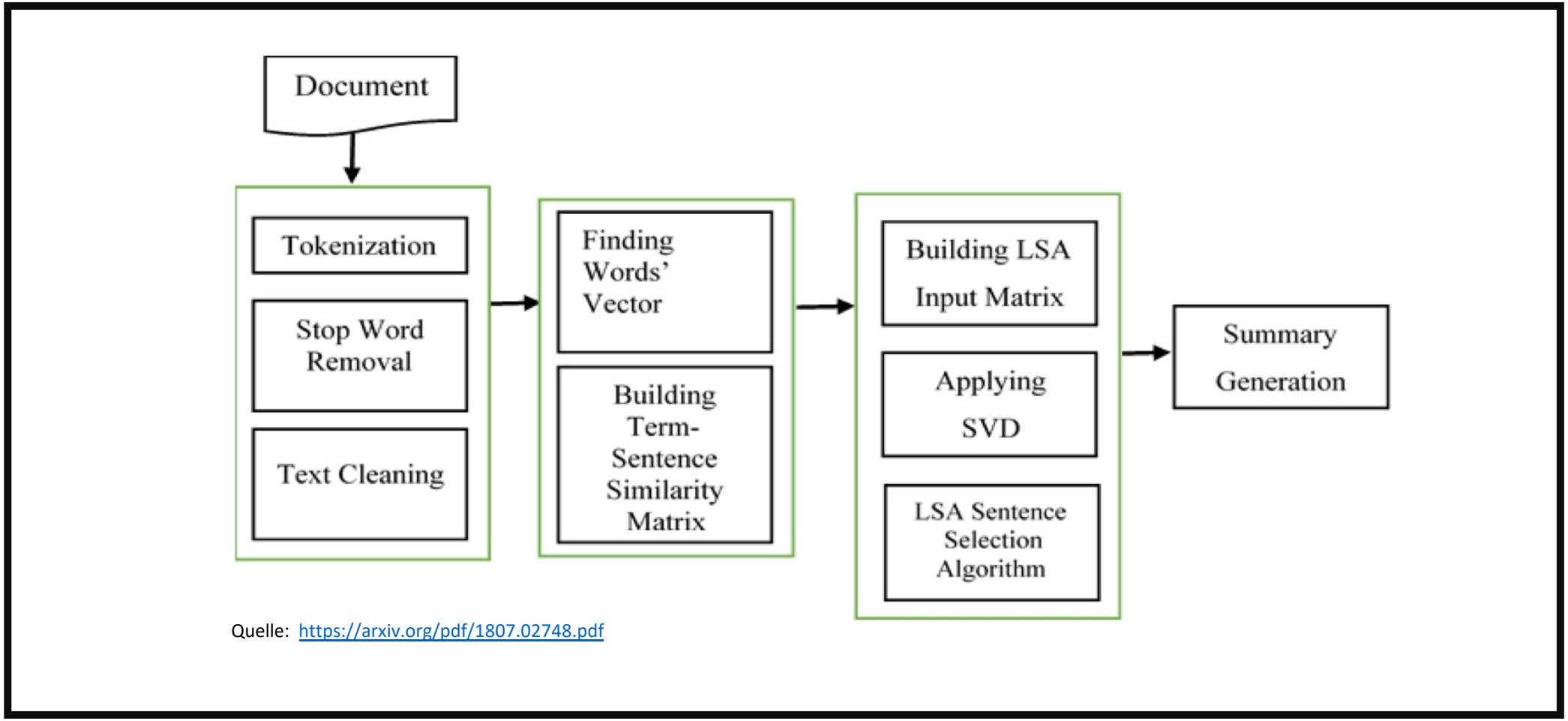


TextRank - Algorithmus



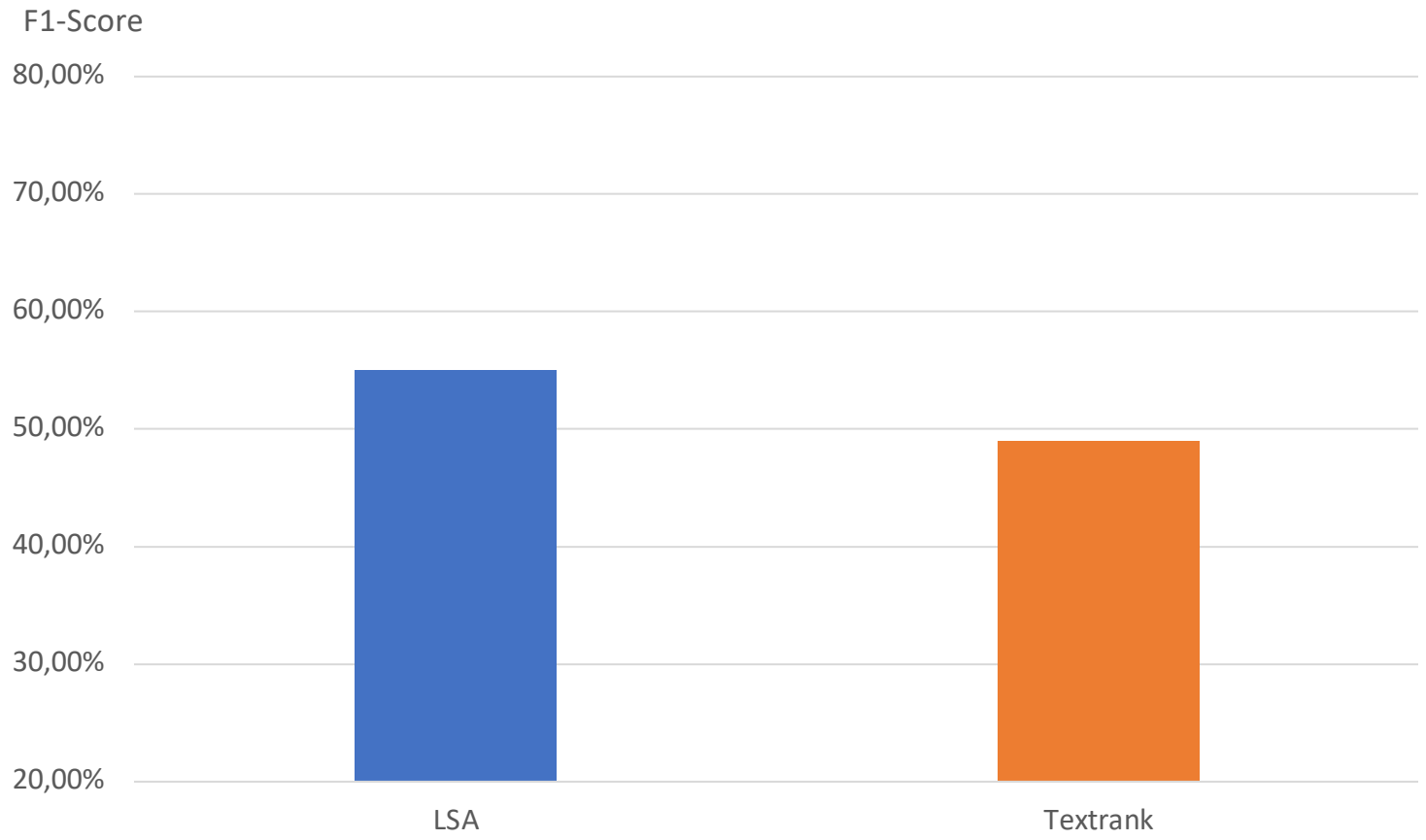
Quelle: https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/

LSA- Algorithmus



Quelle: https://arxiv.org/pdf/1807.02748.pdf

Evaluierung

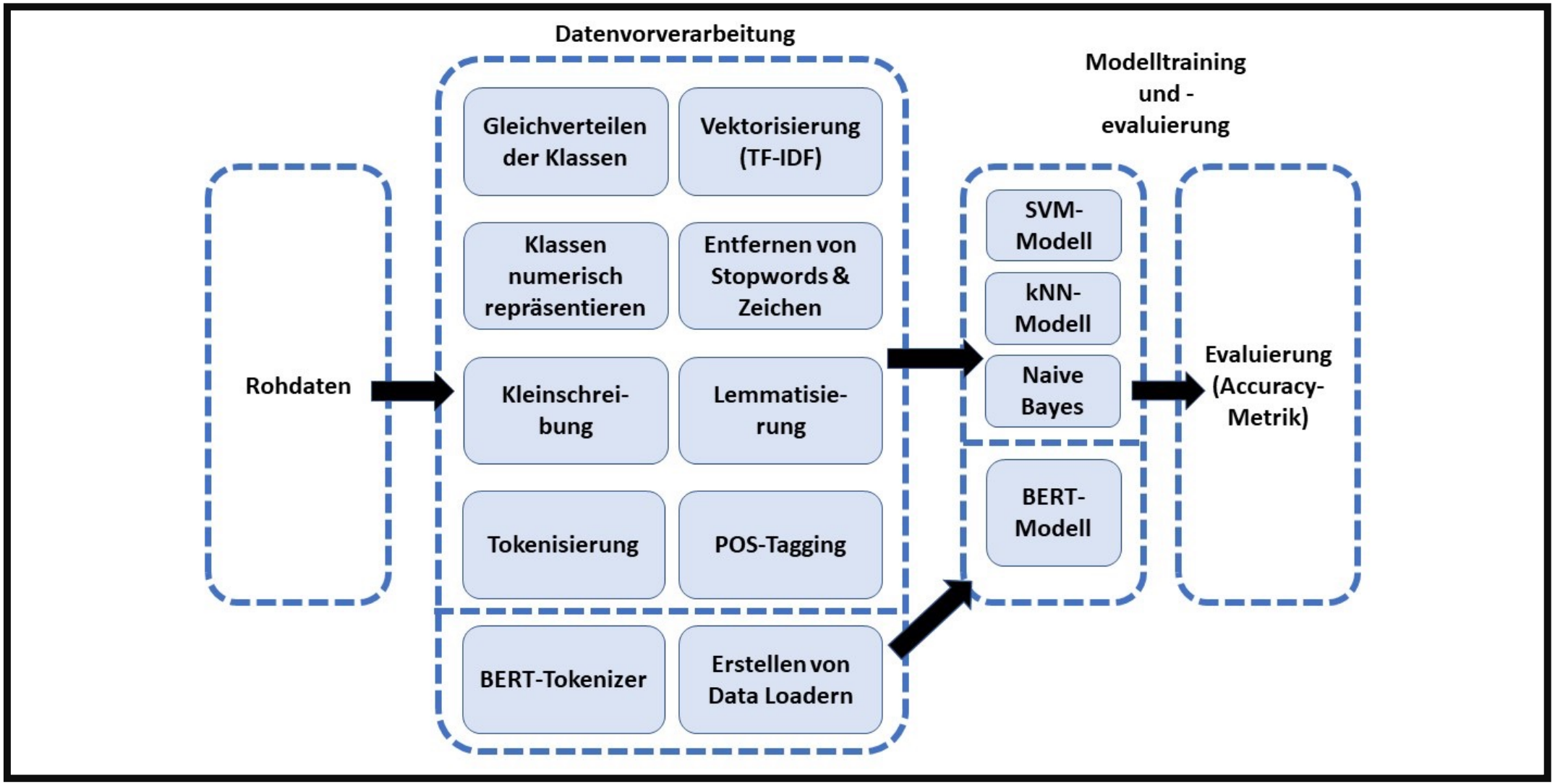


LSA wird ausgewählt

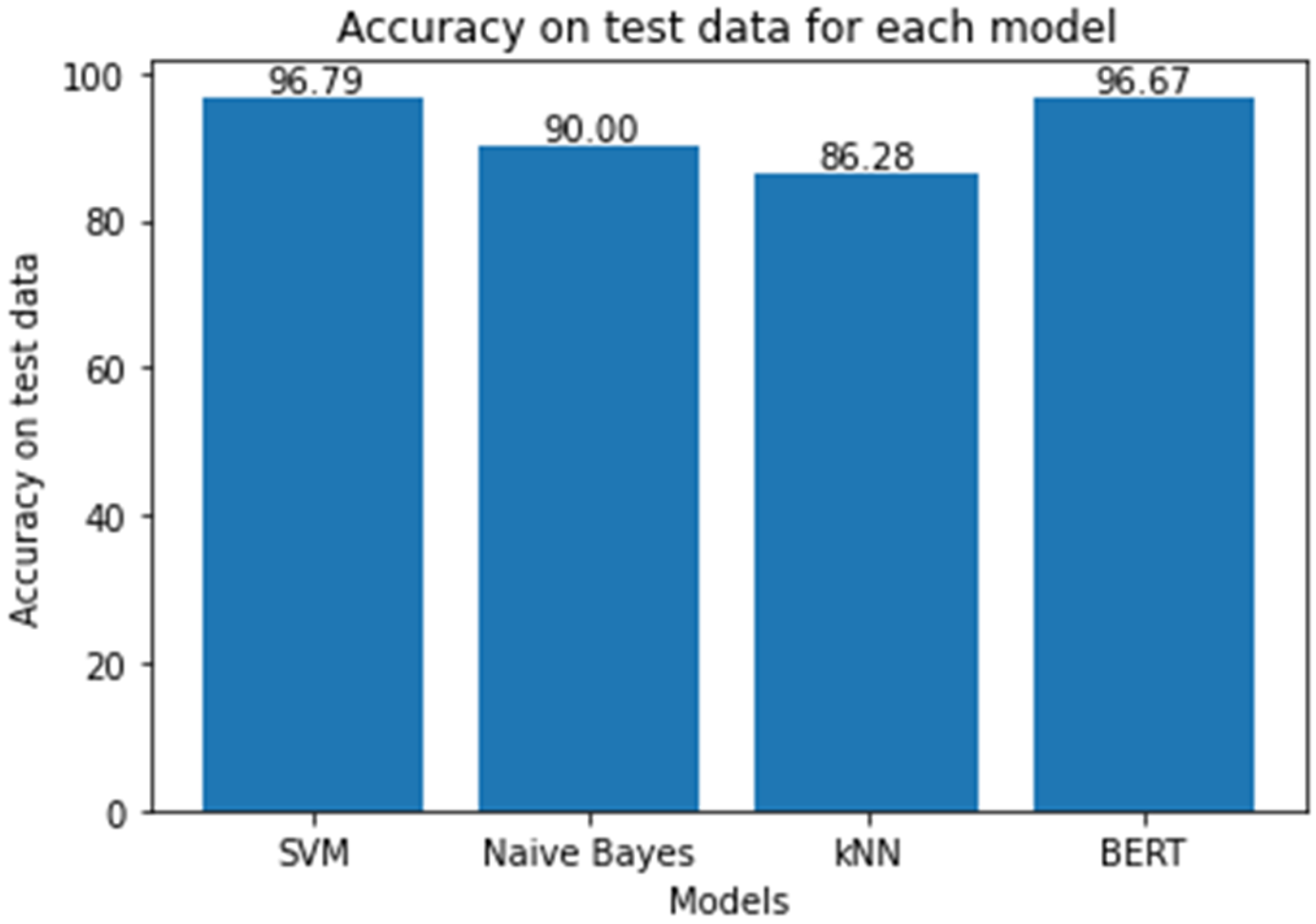
Textzusammenfassung - Erfüllung der Anforderungen

Erfüllt		Noch offen
Kompressionsrate 20%-80%	✓	✗
Englische Sprache	✓	
Zusammenfassung unterschiedlicher Textarten	✓	
Eigenständige Auswahl der Toolchain (Spacy, NLTK, Sumy)	✓	
Extrahierung wichtigster Informationen	✓	

Implementierung Klassifikation



Evaluierung der Textklassifikation

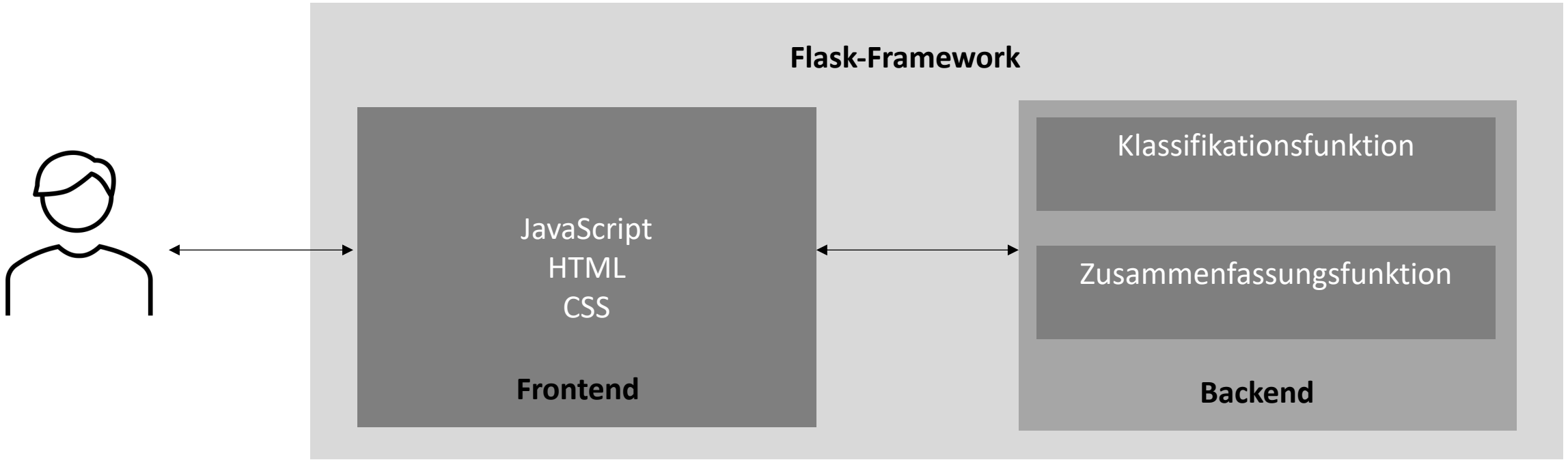


Mit einer Accuracy von **96,79%** auf den Testdaten hat das **SVM-Modell** die besten Ergebnisse erzielt

Klassifikation - Erfüllung der Anforderungen

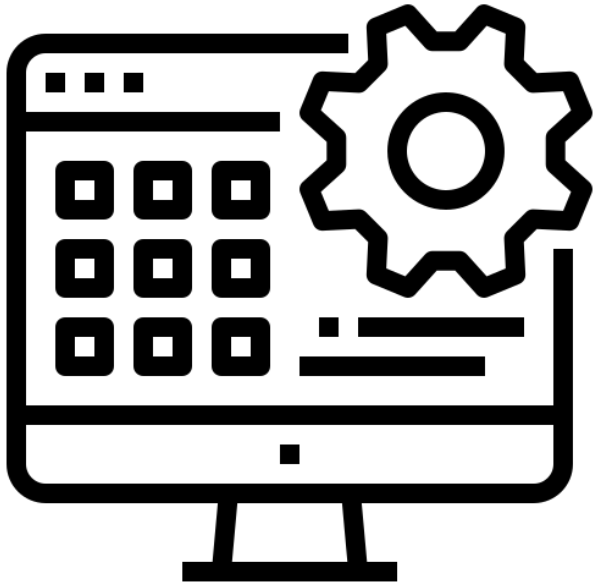
Erfüllt		Noch offen
Präzise Klassifikation in Oberkategorien	✓	✗
Auswahl geeigneter Modelle und Algorithmen	✓	
Eigenständige Auswahl der Toolchain (Spacy, Scit-kit Learn)	✓	
Klassifikation englischer Texte	✓	

Implementierung Webanwendung



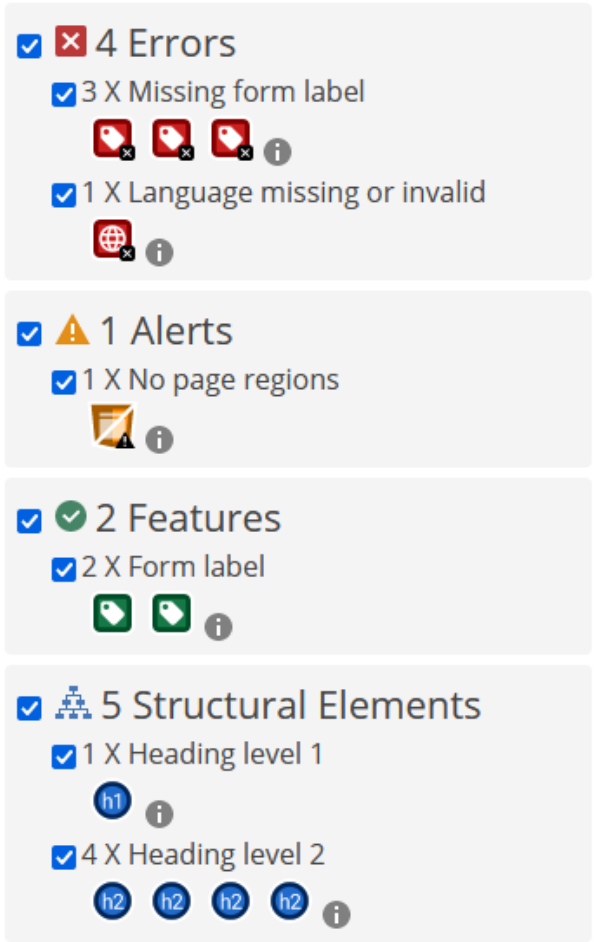
Funktionen Webanwendung

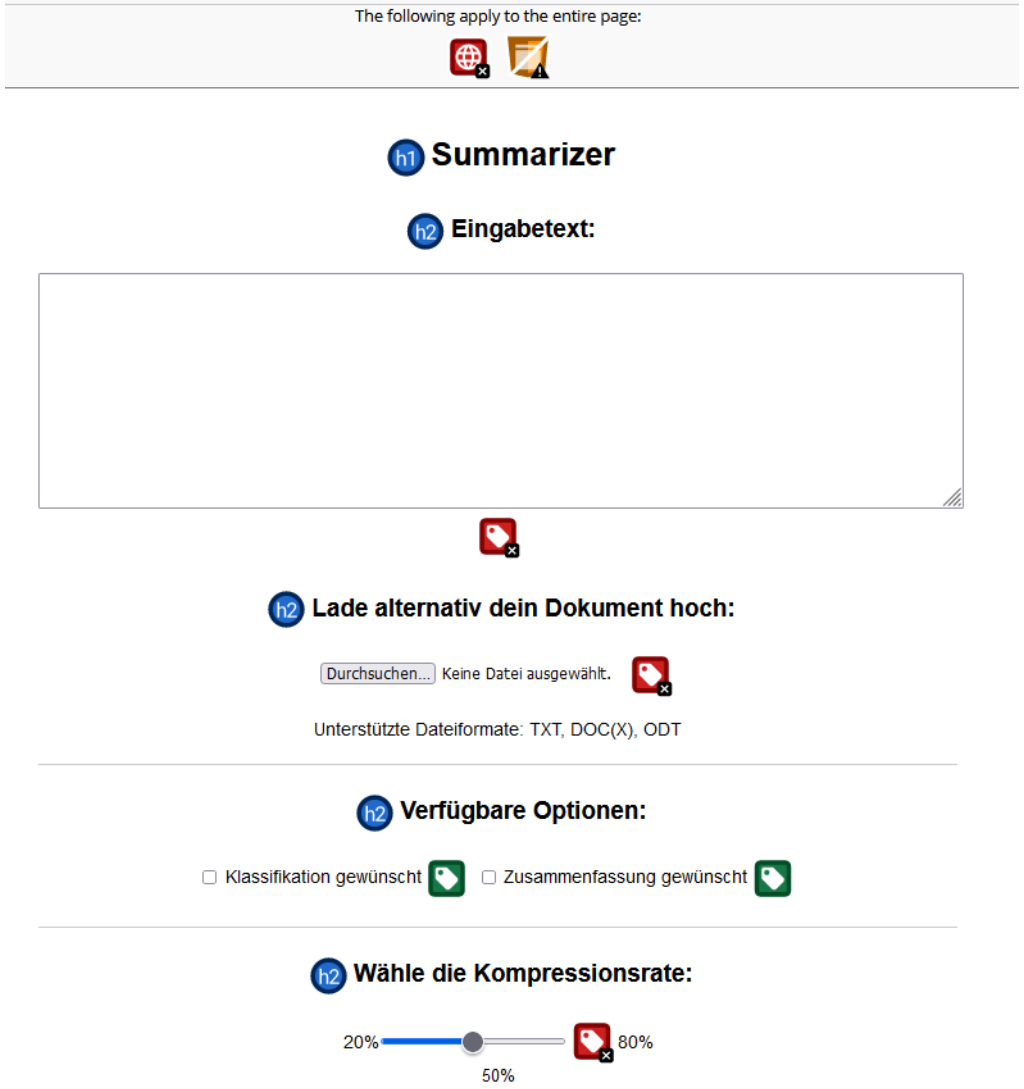
- Texteingabe
- Upload von TXT-,DOC(X)- und ODT-Dateien
- Speech-to-Text
- Schieberegler zur Auswahl der Klassifikationsrate (20%-80%)
- Auswahl der gewünschten Funktionen (Klassifikation und/ oder Zusammenfassung)
- Barrierefreiheit: Superzoom & Kontrast ändern
- Vergleich: Gewünschte und tatsächliche Kompressionsrate

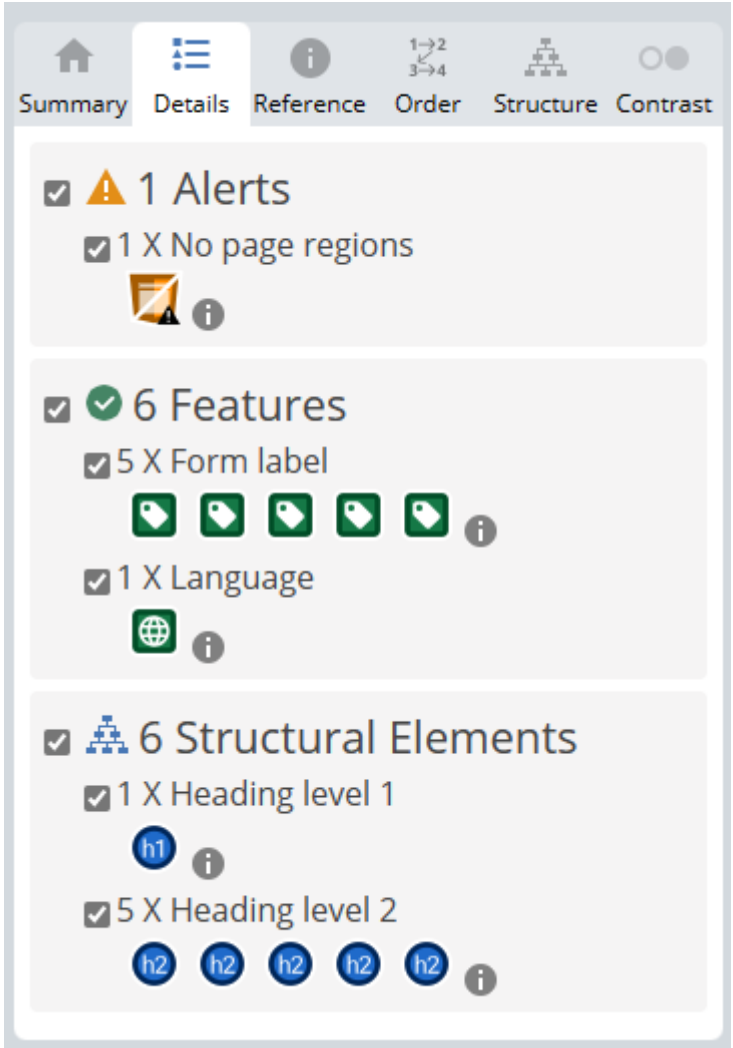


Ergebnisdemo...

Evaluierung Barrierefreiheit - WAVE





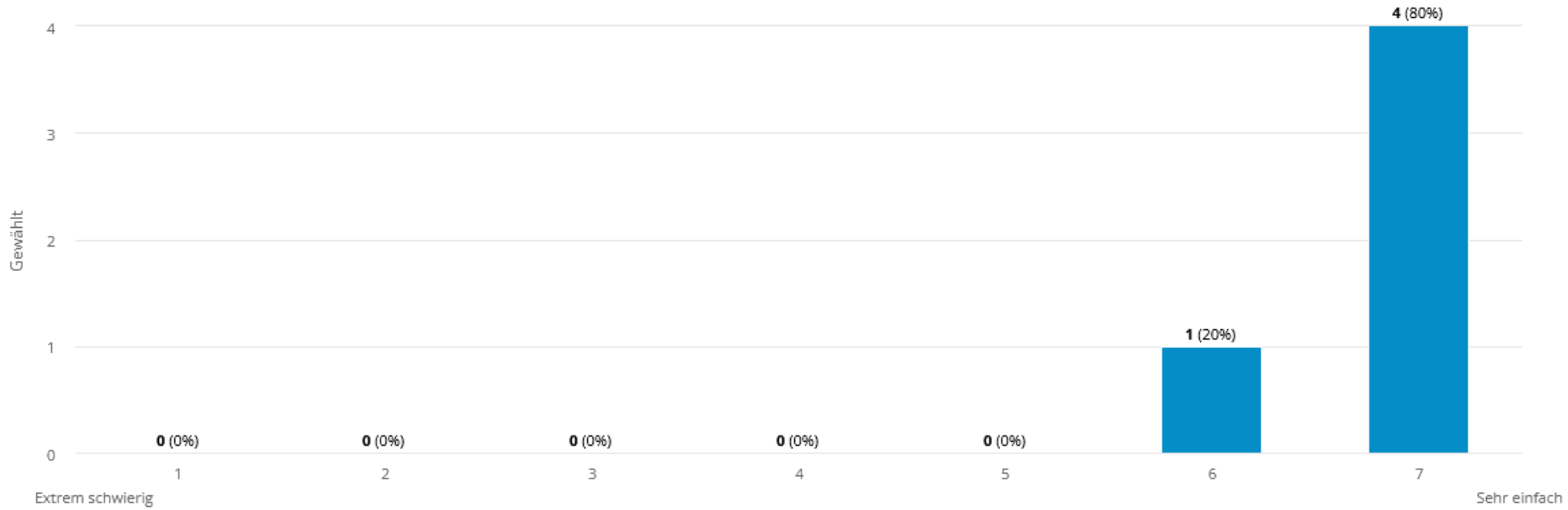


Evaluierung Webanwendung

https://www.umfrageonline.com/user/surveys/1637003/reports

Wie einfach war es für Sie, die Website zu nutzen und die gewünschten Funktionen zu verwenden? Hatten Sie Schwierigkeiten oder fanden Sie die Benutzeroberfläche intuitiv gestaltet?

Anzahl Antworten: 5



Webanwendung - Erfüllung der Anforderungen

Erfüllt

- Texte können in unterschiedlichen Formaten entgegengenommen werden (Textfeld, ODT, DOC(X), TXT) ✓
- Speech-to-Text ✓
- Auswahl einer Kompressionsrate ✓
- Anzeigen der Resultate ✓
- Benutzerfreundlichkeit & Barrierefreiheit (Super Zoom, Kontraste) ✓

Noch offen

- Finale Zusammenfassungsfunktion einbinden ✓
- Evaluierung der Benutzerfreundlichkeit & Barrierefreiheit ✓
- ✗

GANTT Chart Resume

1. Keine Überschreitung des Zeitrahmens
2. Begrenzte Möglichkeiten zur Parallelisierung nach GANTT Chart
3. Langer Zeitrahmen für die Datengrundlage durch Stakeholderanforderung
4. Flüssiges Arbeiten ab dem 26. Mai
5. Ab dem 19.6 Verzögerung bei der Einbindung der Textzusammenfassung

NLP-Tool

WWI20DSA

Amina Uicker-Darwish, Aymane Bouguern, Jan Rüdt, Lukas Bonn

Projektanfang: Mo, 8.5.20

Anzeigewoche: 1

AUFGABE	ZUGEWIESEN AN	FORTSCHRITT	START	
1. Textzusammenfassung				
1.1 Erstellung der Datengrundlage	Gesamtes Projektteam	100%	8.5.23	2
1.2 Erstellen der Referenzusammenfassungen	Amina Uicker-Darwish	100%	27.5.23	9
1.3 Erstellen der eigenen Text-Zusammenfassungsfunktionen	Aymane Bouguern, Amina Uicker-Darwish	100%	27.5.23	1
1.4 Evaluierung der Textzusammenfassungsfunktionen	Aymane Bouguern, Amina Uicker-Darwish	80%	15.6.23	2
2. Textklassifikation				
2.1 Erstellung der Datengrundlage	Gesamtes Projektteam	100%	8.5.23	2
2.2 Durchführen der Datenvorverarbeitung	Lukas Bonn	100%	27.5.23	2
2.3 Modellerstellung	Lukas Bonn, Aymane Bouguern	100%	3.6.23	7
2.4 Modellevaluierung	Lukas Bonn	100%	9.6.23	1
2.5 Erstellen der Klassifikationsfunktion	Lukas Bonn	100%	12.6.23	1
3. Webanwendung				
3.1 Erstellen der Webanwendung	Jan Rüdt	100%	17.5.23	2
3.2 Einbinden der Klassifikationsfunktion	Jan Rüdt	100%	16.6.23	1
3.3 Einbinden der Textzusammenfassungsfunktion	Jan Rüdt	0%	22.6.23	2
3.4 Evaluierung und potenzielle Verbesserung der Benutzerfreundlichk	Jan Rüdt	0%	28.6.23	1

Herausforderungen

- Datenerstellung:
  - Neue Anforderungen (5 Oberkategorien)
  - 5 geeignete Datensätze finden (+ Referenzzusammenfassungen erstellen)
  - Größe der Datensätze
  - Datenvorverarbeitung
- Evaluierung der Zusammenfassungsfunktion
- Einbindung der Zusammenfassungsfunktion in die Webanwendung

Lessons Learned

Bad

- Frühzeitige Abstimmung mit den Stakeholdern bezüglich der Erstellung der Datengrundlage
- Absprachen unter den Teammitgliedern schriftlich festhalten

Good

- Etablierung von Wochenzielen innerhalb der Arbeitspakete