

ANGEWANDTE
COMPUTER- UND
BIOWISSENSCHAFTEN



LEON SCHULZ

SOFTWAREPROJEKT – BUCHKLAPPENTEXTE (COARSE)

SOFTWAREPROJEKT – BUCHKLAPPENTEXTE (COARSE)

AKTUELLER STAND

AKTUELLER STAND

- Zeitplan weiterhin eingehalten
- genauere Untersuchung der letzten Ergebnisse anhand von Konfusionsmatrizen
- Vergleich der Ergebnisse bei verschiedenen Ansätzen
- weiterhin Anomalien behandeln

AKTUELLER STAND

- Performance der einzelnen Klassifikatoren:
 - Random Forest: 79%
 - Radius Neighbor: 52%
 - Gaussian Naive Bayes: 51%
 - Multinomial Naive Bayes: 52%
 - Bernoulli Naive Bayes: 52%
 - Ensemble: 54%

- genauere Untersuchung der Ergebnisse zeigt jedoch Schwachstellen

AKTUELLER STAND

Predicted Classes

Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5369	32	20	91	7	3	5	0
	Ratgeber	430	698	5	18	2	2	3	0
	Kinder/ Jugend	436	13	661	13	1	2	0	0
	Sachbuch	532	16	8	859	4	5	2	0
	Ganzheitl. Bewusstsein	196	18	3	14	304	1	2	0
	Glaube/ Ethik	135	9	1	11	4	219	1	0
	Künste	27	1	0	8	0	0	58	0
	Architektur Garten	38	3	0	2	1	0	0	51

Abb.1: Konfusionsmatrix – Random Forest

AKTUELLER STAND

Predicted Classes

Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5366	42	22	81	8	2	5	1
	Ratgeber	430	692	3	25	3	1	2	0
	Kinder/ Jugend	436	12	661	12	3	2	0	0
	Sachbuch	520	20	3	875	4	2	2	0
	Ganzheitl. Bewusstsein	194	15	5	15	305	3	1	0
	Glaube/ Ethik	139	9	2	10	1	218	1	0
	Künste	25	1	0	9	0	0	59	0
	Architektur Garten	35	5	0	3	1	0	0	51

Abb.2: Konfusionsmatrix – Random Forest ohne Wörterbuch

AKTUELLER STAND

		Predicted Classes							
Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5527	1	0	3	0	0	0	0
	Ratgeber	1156	0	1	2	0	0	0	0
	Kinder/ Jugend	1126	0	0	0	0	0	0	0
	Sachbuch	1419	0	0	7	0	0	0	0
	Ganzheitl. Bewusstsein	538	0	0	0	0	0	0	0
	Glaube/ Ethik	380	0	0	0	0	0	0	0
	Künste	93	0	0	0	0	0	1	0
	Architektur Garten	95	0	0	0	0	0	0	0

Abb.3: Radius Neighbor

		Predicted Classes							
Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5318	108	2	70	0	11	0	0
	Ratgeber	1048	68	0	34	0	1	0	0
	Kinder/ Jugend	1061	47	0	14	0	1	0	0
	Sachbuch	1292	53	1	61	0	9	0	0
	Ganzheitl. Bewusstsein	512	16	0	8	0	0	0	0
	Glaube/ Ethik	353	14	0	7	0	3	0	0
	Künste	84	4	0	5	0	0	0	0
	Architektur Garten	87	8	0	0	0	0	0	0

Abb.4: Multinomial Naive Bayes

AKTUELLER STAND

		Predicted Classes							
Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5021	39	0	340	42	18	19	0
	Ratgeber	939	61	0	85	40	10	0	0
	Kinder/ Jugend	1037	13	0	54	12	2	3	0
	Sachbuch	1101	18	0	243	31	9	7	1
	Ganzheitl. Bewusstsein	429	27	0	43	28	3	2	0
	Glaube/ Ethik	289	12	0	42	19	9	2	0
	Künste	67	1	0	18	1	0	5	0
	Architektur Garten	64	4	0	12	11	2	0	2

Abb.5: Gaussian Naive Bayes

		Predicted Classes							
Actual Classes		Literatur/ Unterhaltung	Ratgeber	Kinder/ Jugend	Sachbuch	Ganzheitl. Bewusstsein	Glaube/ Ethik	Künste	Architektur Garten
	Literatur/ Unterhaltung	5527	0	0	0	0	0	0	0
	Ratgeber	1156	0	0	0	0	0	0	0
	Kinder/ Jugend	1125	0	0	0	0	0	0	0
	Sachbuch	1425	0	0	0	0	0	0	0
	Ganzheitl. Bewusstsein	535	0	0	0	0	0	0	0
	Glaube/ Ethik	376	0	0	0	0	0	0	0
	Künste	93	0	0	0	0	0	0	0
	Architektur Garten	95	0	0	0	0	0	0	0

Abb.6: Bernoulli Naive Bayes

SOFTWAREPROJEKT – BUCHKLAPPENTEXTE (COARSE)

WEITERES VORGEHEN

WEITERES VORGEHEN

- Einfluss der einzelnen Features untersuchen
- Einfluss der Überrepräsentation einzelner Genre in Trainingsdaten (Literatur/Unterhaltung)
- Ensemble - Lösung mit unterschiedlichen Ansätzen (Features, Parameter, Wichtungen)?
- weitere Anomalien behandeln

SOFTWAREPROJEKT – BUCHKLAPPENTEXTE (COARSE)

**VIELEN DANK FÜR IHRE
AUFMERKSAMKEIT**