Automated Tenor Analysis in German Legal Court Rulings

Mario Turic, March 15, 2021, Final Presentation Bachelor Thesis

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Outline

Motivation

- The German court systems is a central parts of the german society
- Growing amount of manual and repetitive labor done by lawyers and judges
- Reduced time for high quality work due to higher quantity of court rulings to analyzed and writen
- Opportunities through the usage and enhancements of software and natural language processing

Goal: Improvement of the current digitalization for the german civil court system through...

- ***Automated legal consequence extraction from tenor texts***
- **Derivation of legal consequences using *court ruling instance trains***

Outline

Research questions

RQ1: Which techniques & concepts, from Natural Language Processing in the context of *german (semi-) structured texts*, can be used to automatically extract *legal consequences* from a given *tenor* ?

RQ2: To what extent can the legal consequences from prior instances be derived by the tenor from the last instance, and is it required to consider other parts of the court rulings as well to make the derivation more accurate ?

Research approach taken – Design Science

- Construction of a text corpus
- Interviews with legal domain experts
- Evaluation of the artifacts/software build

Outline

Related Work

- Urchs et al. tried to classify different legal writing styles of lawyers using binary classification algorithms. They also stated that the publicly available already annotated corpora of German legal texts are rare.

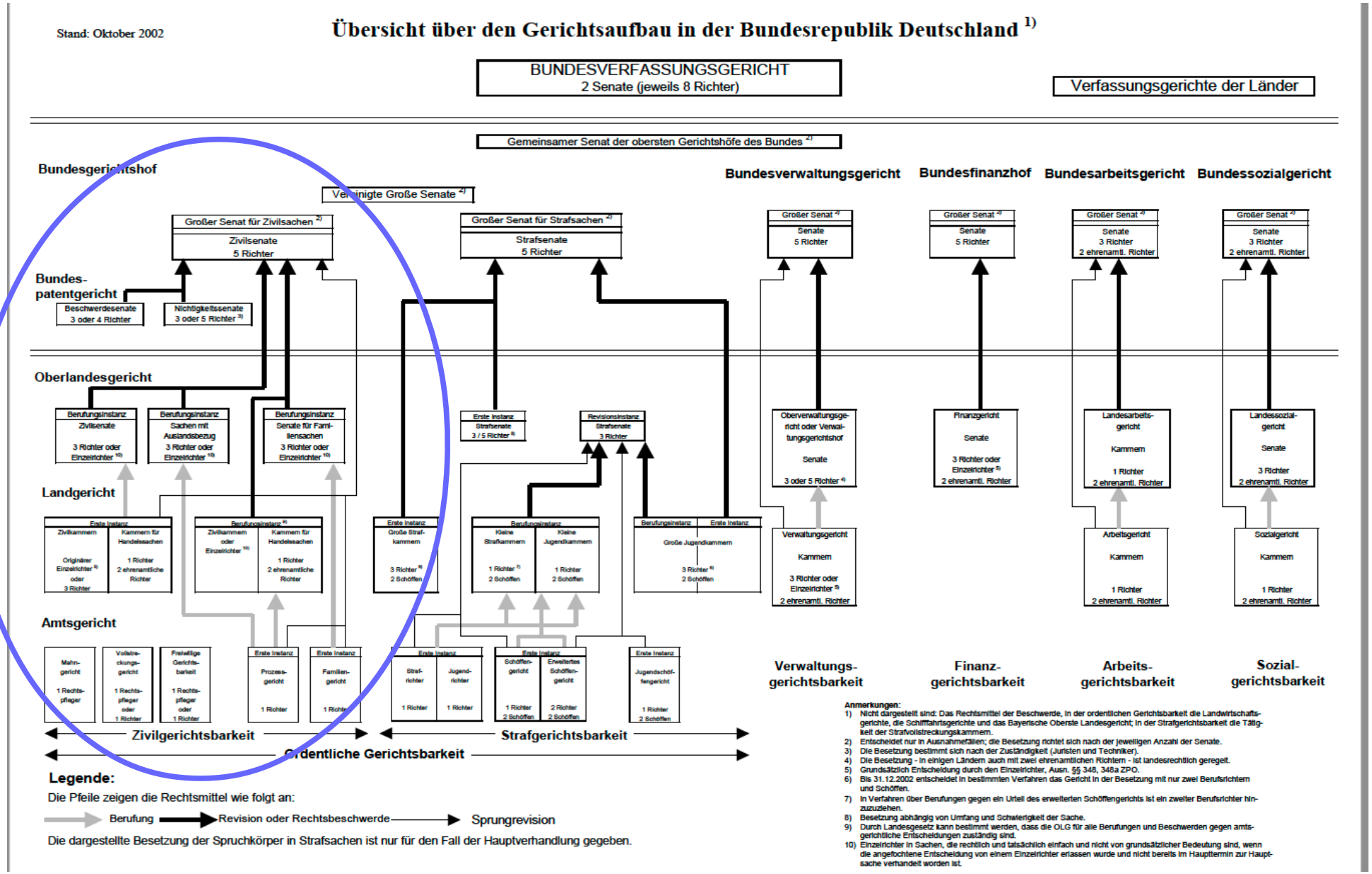
- Leitner et al. established a freely available dataset containing 705 documents were collected and annotated for the task of Named Entity Recognition (NER), whereas 107 court decisions from seven different courts allocated in Germany were collected.

- Ashely et al. examined the question how information can be extracted from already made court rulings and how these information can be used to predict the outcome of new law cases in terms of correct or abstinent decisions made by the classifier.

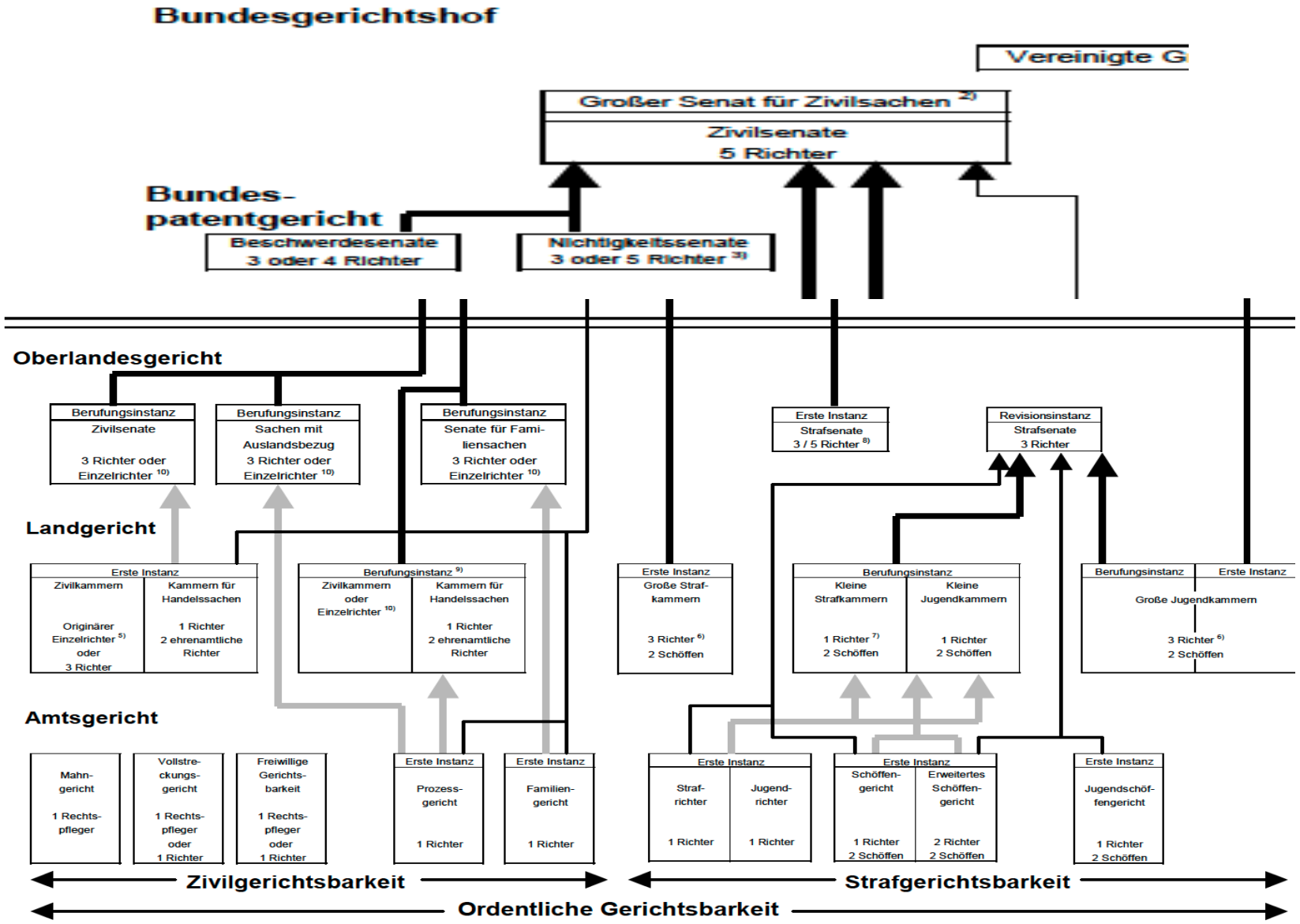
- Pranckevičius et al. presented a research paper comparing naive bayes, randomforest, decision tree and support vector machines performance for the problem of multi class text classification. an interval with minimum of 24.1% and an maximum of 58.5% of average accuracy.

Outline

Domain Overview (I) – German Court System



Domain Overview (II) German civil court system



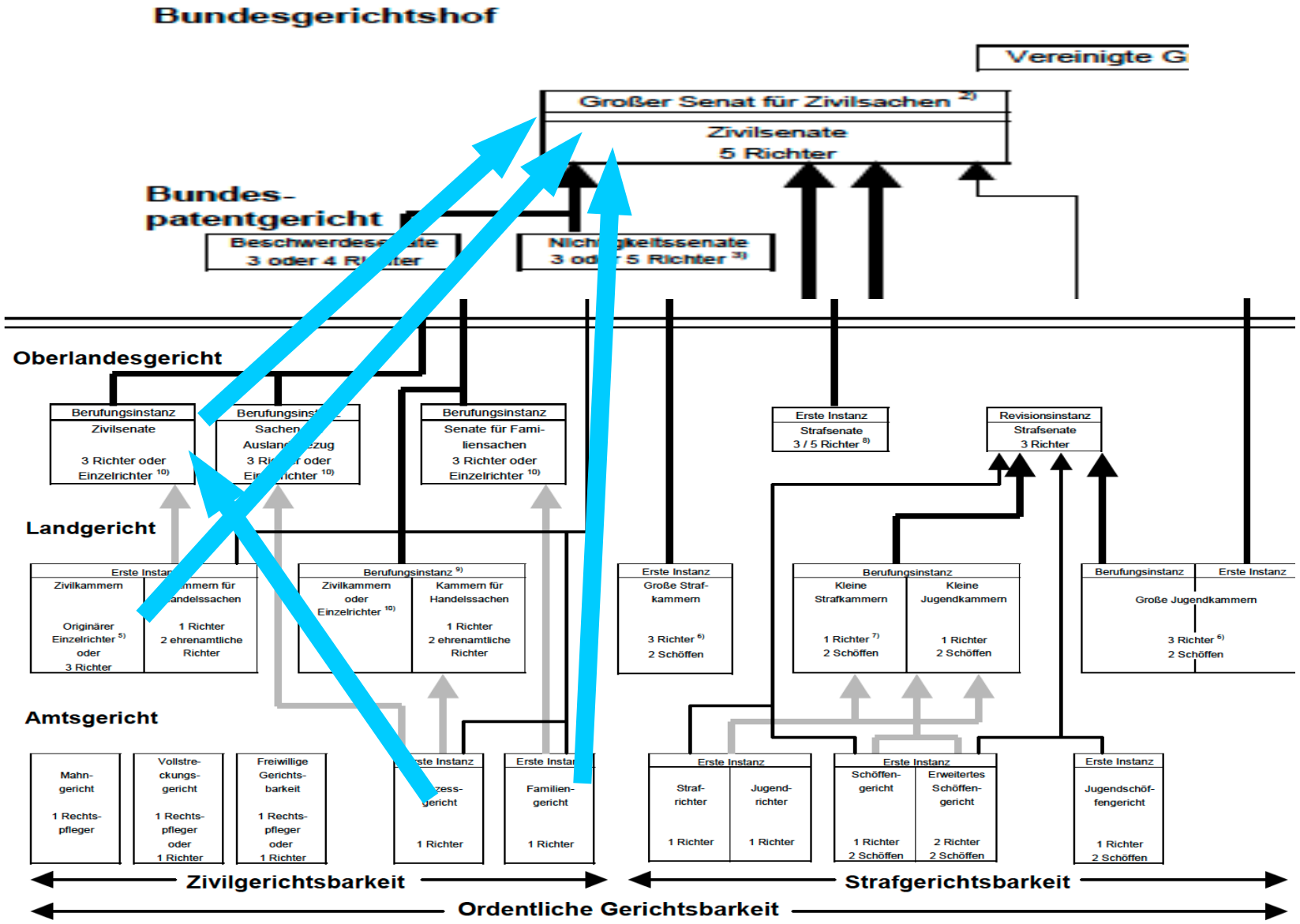
„Court ruling instance train“

1. Instance: AG or LG
2. Instance: LG or OLG
3. Instance: OLG or BGH or BPG

References

- General court system in Germany [BmJV1]

Domain Overview (II) German civil court system



„Court ruling instance train“

1. Instance: AG or LG
2. Instance: LG or OLG
3. Instance: OLG or BGH or BPG

References

- General court system in Germany [BmJV1]

Domain Overview (III) – German court ruling instance train (Theory)

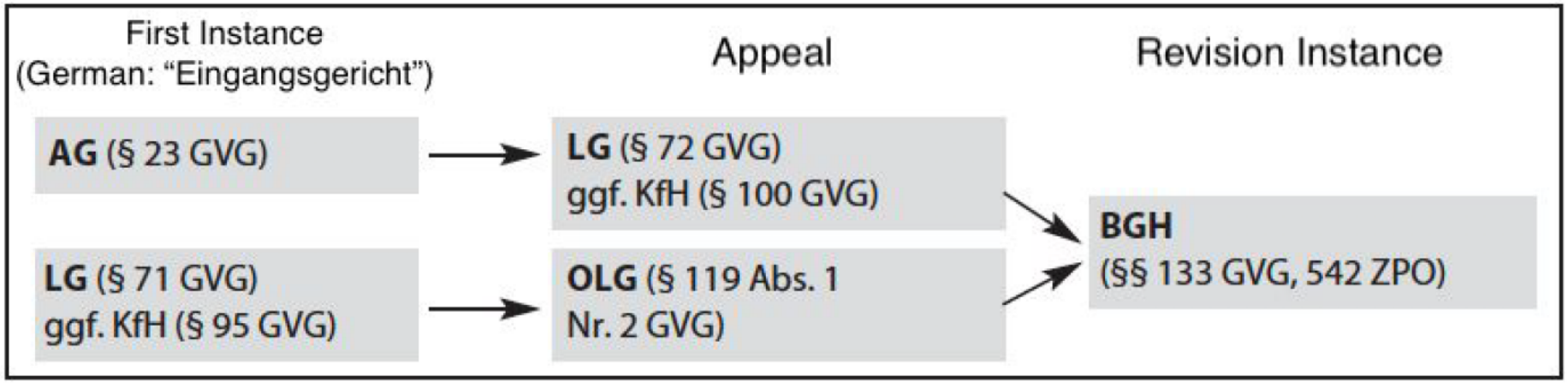


Figure: General Instance Train occurrences
Source: Own illustration based on Mascholleck ZPO, Rn.13; 21. Auflage 2017

Domain Overview (III) – German court ruling instance train (Practice)

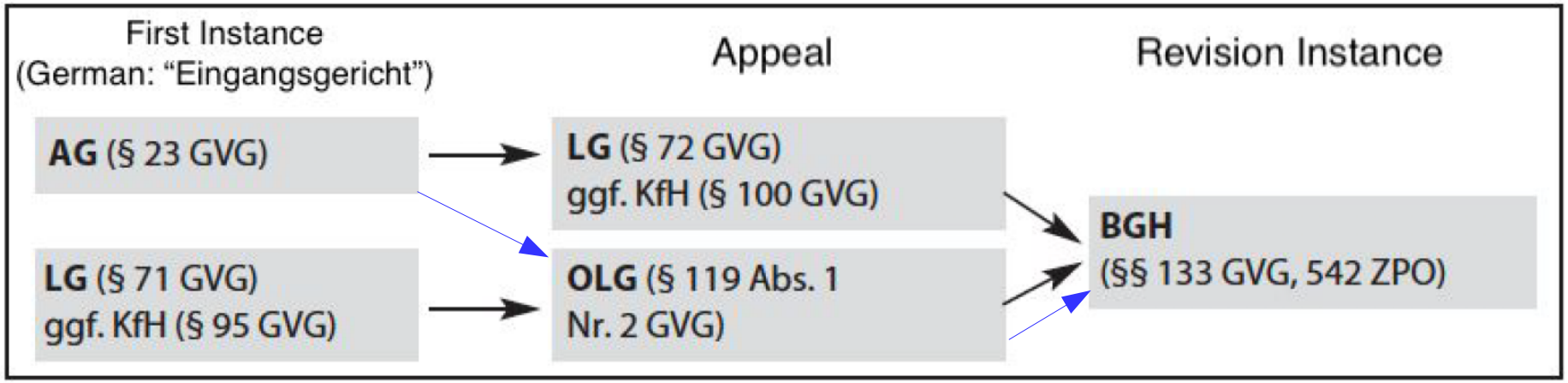
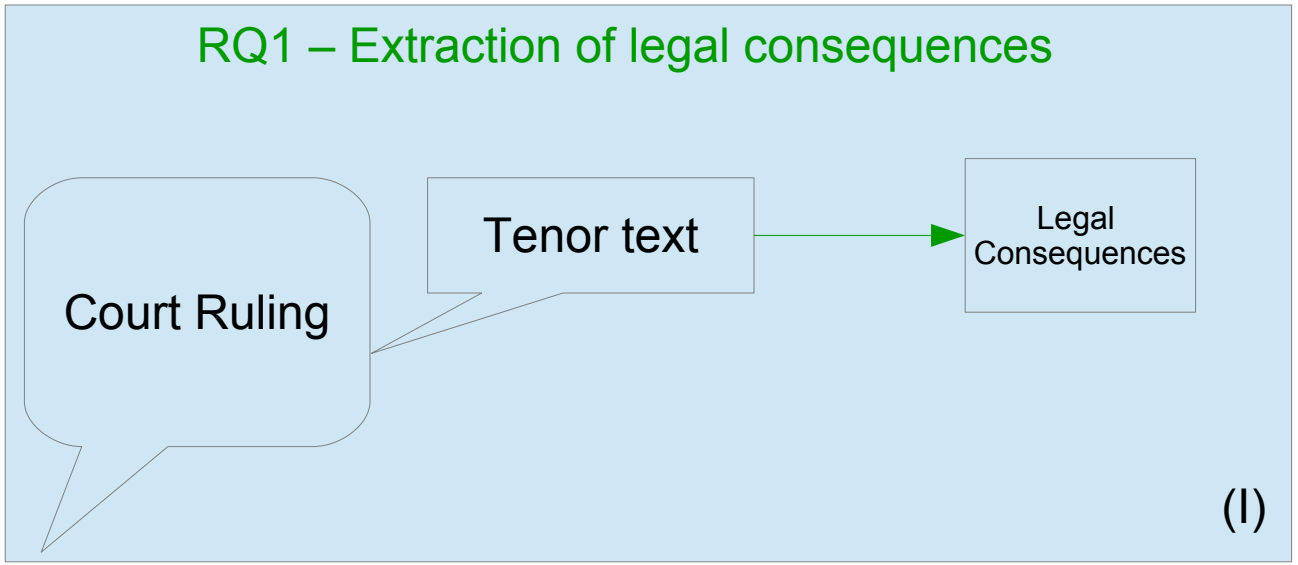
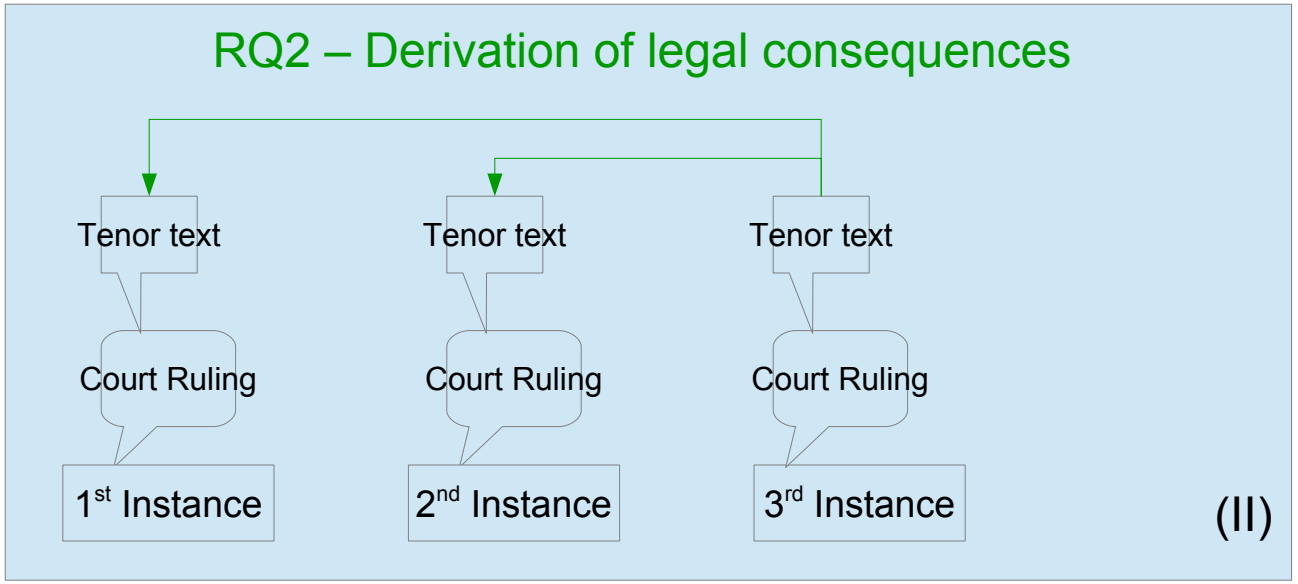


Figure: General Instance Train occurrences
Source: Own illustration based on Mascholleck ZPO, Rn.13; 21. Auflage 2017

Domain overview (V) – Core constructs of interest





Urteilsaufbau			
Urteilskopf (§ 313 I ZPO)			
Urteilsformel (Tenor) (§ 313 I Nr. 4 ZPO) zu			
1. Hauptsache	2. Kosten des Rechtsstreits	3. Vorläufige Vollstreckbarkeit	4. Prozessualen Nebenentscheidungen
Tatbestand (§ 313 I Nr. 5, II ZPO)			
Entscheidungsgründe (§ 313 I Nr. 6, III ZPO)			

References

- General court ruling structure [ZPO] [Hofmann]
- General tenor structure [Kurpat]

Domain overview (VI) – Interviews with german civil court judges

Interview Preparation:

- An amount of 119 civil court judges contact data was prepared and contacted
- 5 particular judges responded and participated in the interviews
- Guided interviews with prepared questionaire

Interview Execution:

- Phone/Videocalls with the judges
- Recorded the interviews
- Used the recordings to postprocess the insights

Interview Results:

- The questions were answered by judges aged 28 to 56.
- The affinity towards technology and software was generally in the middle (3.5) on a numeric scale from 1 (low) to 5 (high)
- Read approximately two to three court rulings completely throughout a work week and
- Scan more than 20 court rulings on the average in a work week
- Tenor and the legal consequences of the cases are not in their main focus.
- Each of the interviews can be found in full length in the appendix of the bachelor thesis.

Domain overview (VII) – Tenor text corpus

- The court ruling data were collected by crawler, manually and mostly in .pdf file format
- 3030 tenor texts from court ruling documents were labeled for RQ1:

696	BGH court rulings (PDF)
1366	OLG court rulings (PDF)
434	LG court rulings (PDF)
442	AG court rulings (PDF)
92	KG court rulings (PDF)

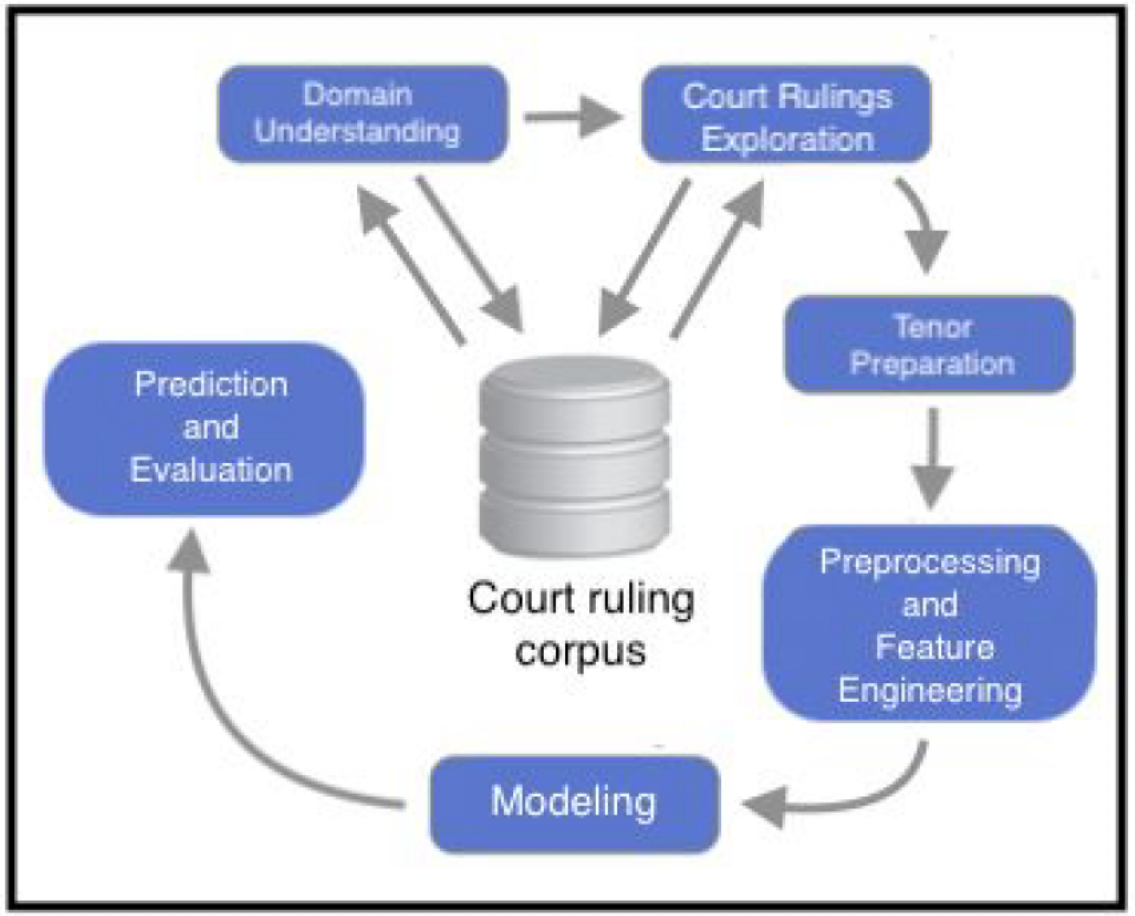
- 300 complete instance trains were collected and labeled for RQ2:

242	“BGH - OLG - LG” instance trains
48	“BGH - LG - AG” instance trains
0	“OLG - LG - AG” instance trains
10	“BGH - OLG - AG” instance trains

- Manual labeling using the labeling manual constructed by Ingo, lawyer and me
- Tenor text data was stored insight an .xslx file(s) and parsed (shown in the following slides)

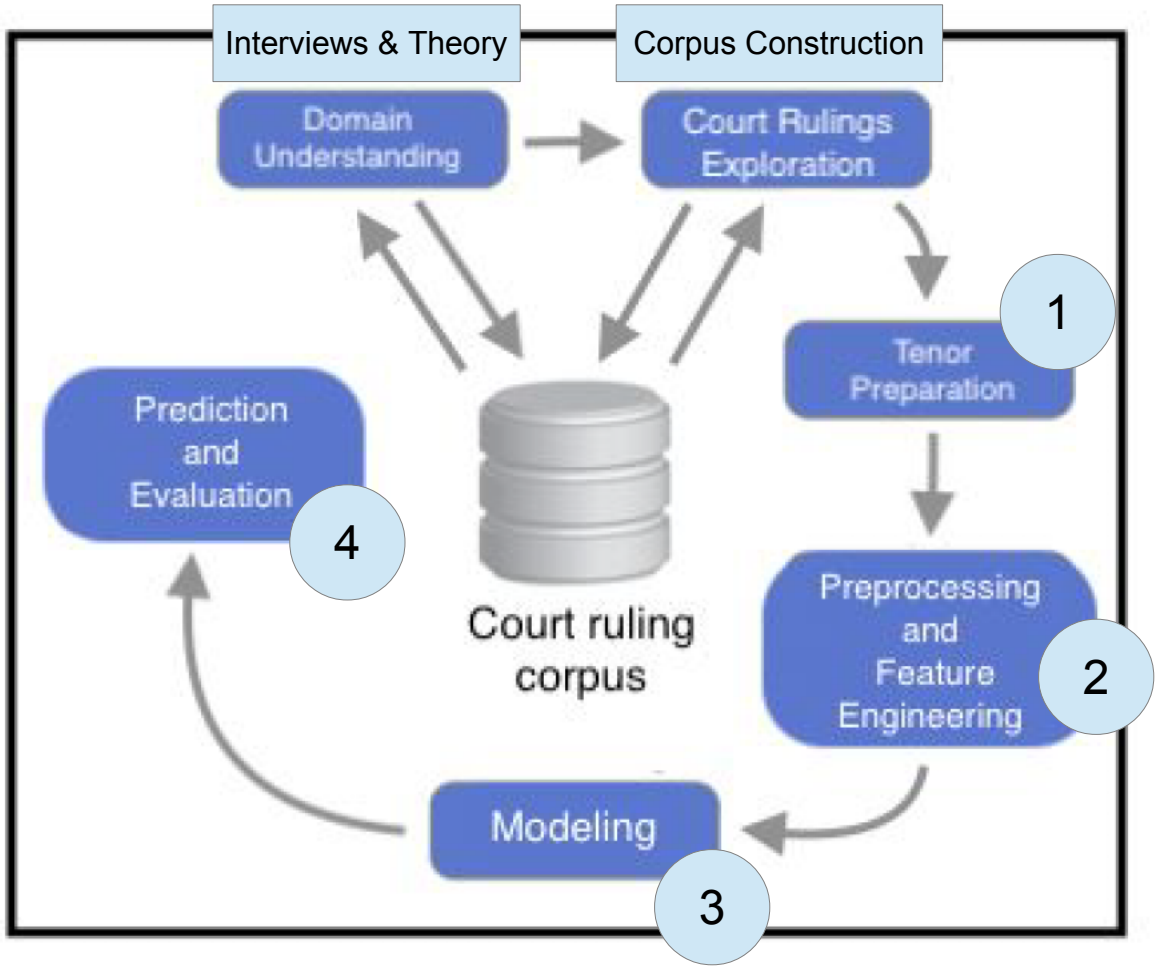
Outline

Curt Rouling Pipeline



Inspiration Source: Wirth, Rüdiger, and Jochen Hipp - CRISP-DM: Towards a standard process model for data mining

Curt Rouling Pipeline



Tenor Preparation – Labeled Tenor text data for RQ1

A.1. Annotation manual

ANNOTIERUNGSKLASSEN

1. Hauptausspruch / Erfolgsaussicht:

E = Erfolg (Stattgabe der Klage, Aufhebung des Urteils...)

N = Nichterfolg (Abweisung als unbegründet (erste Instanz), Zurückweisung als unbegründet (Revision, Rechtsbeschwerde), Verwerfung als unzulässig ...)-> hier kann man weiter differenzieren

T = Teilerfolg (Teilabweisung)

V = Vorlage an EuGH

K = keine Entscheidung in diesem Verfahren (Aussetzung des Verfahrens)

NT (Kombinationen) = Nichterfolg und Teilerfolg (bei von Kläger und Beklagtem eingelegten Rechtsmitteln)

2. Kostenentscheidung

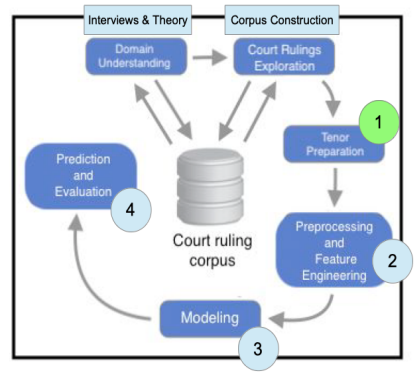
K = Kläger (Antragssteller) trägt Kosten des Verfahrens

B = Beklagter (oder Antragsgegner) trägt Kosten des Verfahrens

T = Kostenteilung

A = Kostenaufhebung

KK = keine Kostenentscheidung (es steht gar nichts dazu im Tenor ODER es bleibt einem anderen Verfahren vorbehalten)



3. Vollstreckbarkeit

VV = vorläufig vollstreckbar nach § 708 ZPO

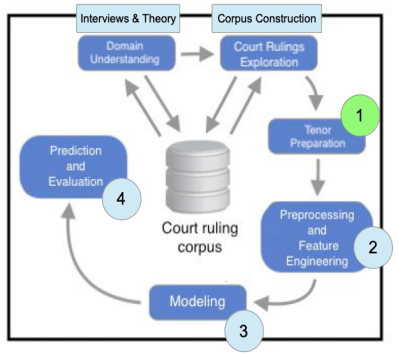
VS = vollstreckbar gegen Sicherheitsleistung nach § 709 ZPO

VA = vollstreckbar mit Abwendungsmöglichkeit nach §§ 710, 711 S. 3 und 712

KE = keine Entscheidung

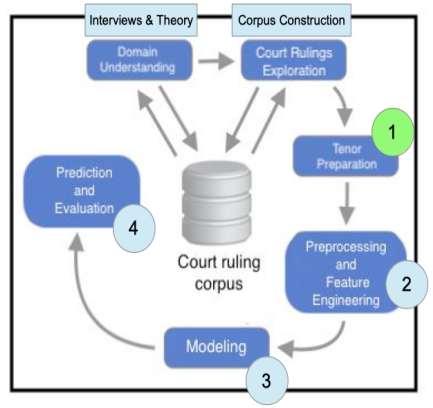
Tenor Preparation – Labeled Tenor text data for RQ1

7	Total Amount of tenor data	Court Ruling File Number	Tenor Text	Main Decision	Cost Decision	Enforceability
8	1	33 O 22637/14	I. Die Beklagten	E	B	VS
9	2	51 O 12/02	I. Die Klage wird	N	K	VS
10	3	9 O 77/14	Die Klage wird	N	K	VS
11	4	11 O 11/14 KfH	1. Die	K	KK	VV
12	5	26 C 76/13	Die Beklagten	E	B	VV
212	205	15 O 50/15	Die Klage wird	N	K	VS
213	206	2 HKO 2542/14, 2 HK O	1. Die Klage wird	N	K	VA
214	207	34 O 24/12	Die Klage wird	N	K	VS
215	208	1 O 147/11	Die Klage wird	N	K	VS
216	209	15 O 314/14	Die Klage wird	N	K	VS
587	580	6 U 174/10	Die Berufung der	N	B	VA
588	581	6 U 139/10	Die Berufung der	N	B	VA
1070	1063	1 S 50/13	Die Berufung des	N	B	VV
1071	1064	21 S 68/13	1. Auf die Berufung	T	B	VV
1072	1065	51 O 25/13	Die Klage wird	N	K	VS
2005	1998	IX ZR 128/12	Die Revision gegen	N	B	KE
2006	1999	IX ZR 127/11	Auf die	E	T	KE
2007	2000	IX ZR 118/12	Die Revision gegen	N	K	KE
2008	2001	IX ZR 103/13	Auf die Revision	E	KK	KE
2009	2002	IX ZR 80/13	Die Revision gegen	N	K	KE
2763	2756	I-6 U 160/13, 6 U 160/13	Die Berufung der	N	B	VA
2764	2757	15 U 61/09	Die Berufung der	N	B	VA
2765	2758	7 U 49/09	Auf die Berufung	T	B	VA
3105	3098	I-1 U 131/12, 1 U 131/12	Auf die Berufung	T	B	VA
3106	3099	10 U 1933/10	I. Auf die Berufung	N	K	VA
3107	3100	3 U 70/12	1. Die Berufung des	N	K	VA
3108	3101	4 U 218/12	I. Auf die Berufung	N	KK	KE
3109						



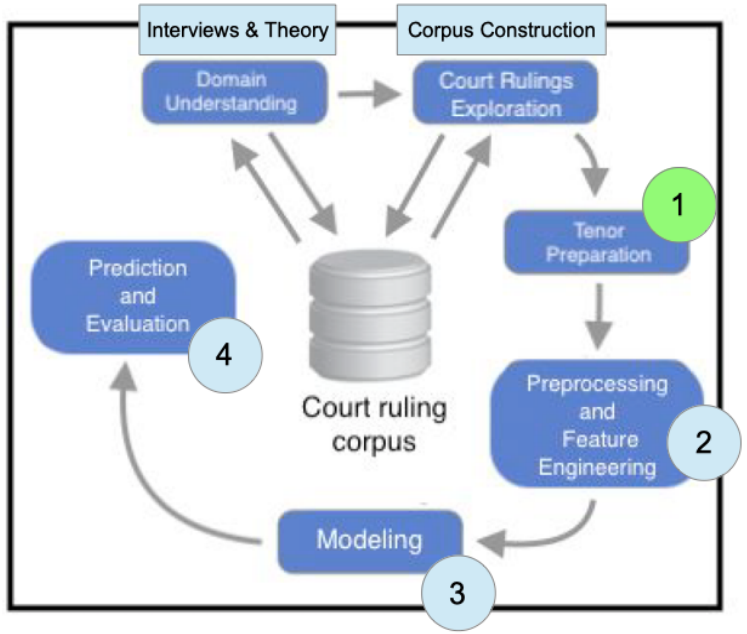
Tenor Preparation – Labeled Instance train data for RQ2

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		tenor_text	tenor_	tenor_	tenor_vollstreck												
2		Annotierung von 10 Ten	Für insgesamt 500 Tenore ca. 15h Zeit benötigt														
3																	
4		Laut "Die Urteilsformel" Tenor	nach Annotierungsklassen von Ingo														
5																	
6																	
7	Revisions Urteil Aktenzeichen	Revisions Tenor Text	Revisic	Revisic	Revisic	ERSTER E	Berufungs Urteil Aktenzeichen	Appeal Tenor Text	Appeal -	Appeal -	Appeal -	Erste Instanz Urteil Ak	First Instance T	First instance -	First instance -	First Instance -	Data source file
8	I ZR 138/16	Auf die Revision der	E	KK	KE		29 U 3500/15	I. Die Berufung der	N	B	VA	33 O 22637/14	I. Die	E	B	VS	erste_instanz_batch01
9	I ZR 18/04	Auf die Revision des	E	KK	KE		6 U 161/02	Die Berufung des	N	K	VA	51 O 12/02	I. Die Klage	N	K	VS	
10	IV ZR 448/14	Auf die Revision der	T	T	KE		I-20 U 110/14, 20 U 110/14*	Auf die Berufung	T	T	VA	9 O 77/14	Die Klage	N	K	VS	
11	IX ZR 232/17	Die Revision gegen das	N	B	KE		12 U 8/17	1. Die Berufungen	T	B	VA	11 O 11/14 KfH	1. Die	K	KK	VV	
12	VIII ZR 175/14	Die Revision des	N	B	KE		25 S 188/13	Die Berufung der	T	B	VA	26 C 76/13	Die Beklagten	E	B	VV	
13	KZR 65/12	Die Revision gegen das	N	K	KE		5 U 229/05	Die Berufungen der	N	K	VA	3/9 O 98/03, 3-9 O	Die Klage	N	K	VS	
14	I ZR 80/12	Die Revision der	T	KK	KE		23 S 343/13	Unter	K	B	VA	34 C 154/13	1. Der	E	B	KE	
15	I ZR 107/14	Auf die Revision der	E	B	KE		16 U (Kart) 22/12	Die Berufung der	N	K	VA	14 O Kart 83/10	Die Klage	N	K	VS	
16	V ZR 169/14	Auf die Rechtsmittel	E	B	KE		5 U 87/09	Die Berufung der	T	B	VA	310 O 93/08	1. Den	E	B	VS	
17	II ZR 192/18	Auf die Revision der	T	K	KE		I-6 U 187/13, 6 U 187/13	Die Berufung der	N	K	VA	14 O 44/13	Die Klage	N	K	VS	
18	I ZR 8/13	Die Revision der	T	T	KE		5 S 297/12	1. Die Berufung der	N	K	VV	36 C 124/12 (12)	1. Die Klage	N	K	VA	
19	III ZR 196/17	Auf die Revision der	T	K	KE		4 U 120/17*	Auf die Berufung	T	T	VA	2-21 O 2/16, 2/21 O	Die Klage	N	K	VS	
20	VIII ZR 266/14	Die Revision der	N	K	KE		I-6 U 182/16, 6 U 182/16	Auf die Berufung	E	B	VA	2-6 O 428/10, 2/6 O	I. Den	E	B	VS	
21	IX ZR 79/18	Auf die Revision des	E	KK	KE		18 S 413/13	1. Die Berufung der	N	K	VA	26 O 151/16	Die Klage	N	K	VA	
22	IX ZR 79/18	Auf die Revision des	E	KK	KE		3 U 49/16	1. Auf die Berufung	T	B	VA	237 C 302/13	1. Die Klage	N	K	VA	
23	II ZR 3/04	Auf die Revision des	E	KK	KE		3 U 49/16	1. Auf die Berufung	T	B	VA	13 O 171/15	1. Das	K	B	VS	
24	I ZR 7/14	Die Revision gegen das	N	B	KE		6 U 56/03	Die Berufung des	T	B	VA	13 O 171/15	1. Das	K	B	VS	
25	XI ZR 174/13	Auf die Rechtsmittel	E	B	KE		I-6 U 96/13, 6 U 96/13	Auf die Berufung	T	T	VA	4 O 177/01	Der Beklagte	E	B	VS	
26	I ZR 19/14	Die Revision gegen das	N	B	KE		3 U 229/12	1. Die Berufung des	N	K	VA	14 O 277/12*	Die Beklagte	T	T	VA	
27	XI ZR 790/16	Die Revision der	N	B	KE		I-6 U 205/12, 6 U 205/12	I. Auf die Berufung	T	T	VA	1 O 91/12	1. Die Klage	N	K	VA	
28	I ZR 108/09	Die Klägerin wird	K	KK	KE		I-6 U 56/16, 6 U 56/16	Auf die Berufung	E	B	VA	28 O 306/11	1. Der	E	B	VS	
29	XII ZR 77/06	Auf die Revision des	T	KK	KE		I-20 U 87/08, 20 U 87/08	Auf die Berufung	E	B	VA	12 O 210/15	Die Klage	N	K	VS	
30	XI ZR 78/08	Die Revision der	N	B	KE		29 U 13/03	Auf die Berufung	T	KK	VA	37 O 51/07	Die Klage	N	K	VS	
31	I ZR 69/08	Die Revision gegen das	N	K	KE		7 U 71/07	Die Berufung der	N	B	VA	8 O 188/00	Die Klage	N	K	VS	
32	XII ZB 447/13	Auf die	E	KK	KE		2 U 319/07	Die Berufung der	N	K	VA	13 O 370/06	1. Die	E	B	VA	
33	I ZR 74/12	Auf die Revision der	T	K	KE		2 U 96/07	Die Berufung des	N	K	VA	3 O 1108/05	Die Klage	N	K	VA	
34	VI ZR 76/17	Auf die Revision der	T	K	KE		I-6 U 67/11, 6 U 67/11	1. Die Berufung der	N	B	VA	11 O 165/07	1. Die Klage	N	K	VS	
35	I ZR 154/15	Die Revision gegen das	N	K	KE		15 U 88/16	Die Berufung der	N	B	VS	28 O 716/10*	Die Beklagten	E	B	VA	
36	IX ZR 125/17	Auf die Rechtsmittel	T	KK	KE		9 S 433/14, 9 S 433/14 (59)	1. Die Berufung der	N	K	VA	28 O 379/15	1. Die	E	B	VA	
37	XI ZR 51/10	Auf die Revision der	E	KK	KE		19 U 165/15*	Auf die Berufung	T	T	VA	2 O 273/10	Der Beklagte	E	B	VS	
38	XI ZR 182/10	Die Revision der	N	K	KE		I-15 U 230/09	Die Berufung der	N	K	VA	1 O 310/08	Die Klage	N	K	VS	
39	I ZR 208/15	Auf die Revision der	T	T	KE		13 U 117/09	Auf die Berufung	N	K	VA	325 O 22/09	I. Die	T	B	VS	
40	IX ZR 204/12	Auf die Revision des	E	KK	KE		I-15 U 129/14, 15 U 129/14	I. Die Berufung der	N	B	VA	7 O 12/14	Die Beklagte	T	T	VS	
41	VIII ZR 200/05	Die Revision der	T	T	KE		8 U 55/11	Die Berufung des	N	K	VA	2 O 419/10	Die Klage	N	K	VS	



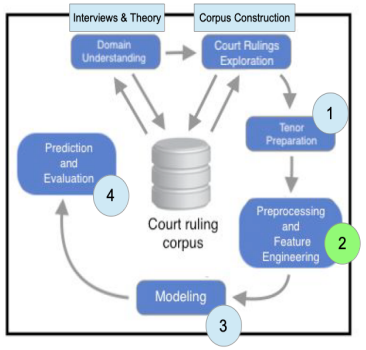
Curt Rouling Pipeline – Tenor Preparation

- Checked crawled court rulings, manuall court ruling search using Juris.de
- Extracted the tenor text data into google spreadsheets
- Labeling of the tenor text data using the labeling manual (constructed by Ingo & me using civil law literature)



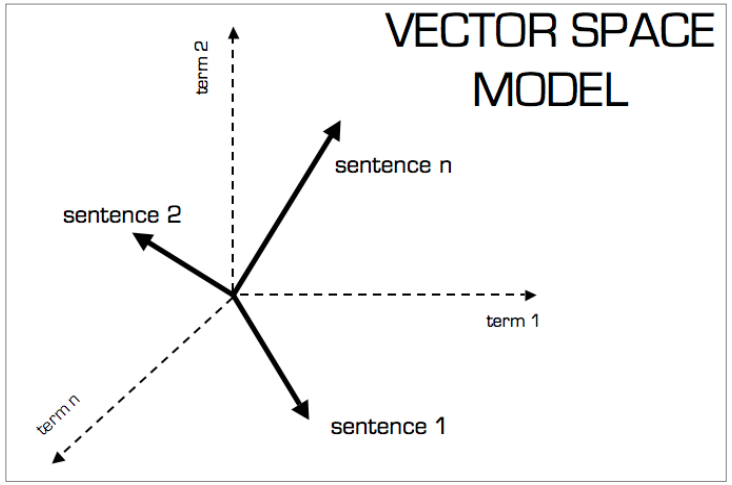
Curt Rouling Pipeline – Preprocessing & Feature Engineering

- Removal of non ASCII characters and etc
- Regular expression and german stop word removal using NLTK corpus for german words
- SnowBall stemmer used for stemming of German words
- Frequency based feature creation (TF-IDF) → TF-IDF Features

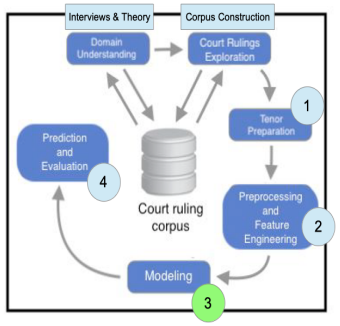


$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents



Curt Rouling Pipeline – Modelling



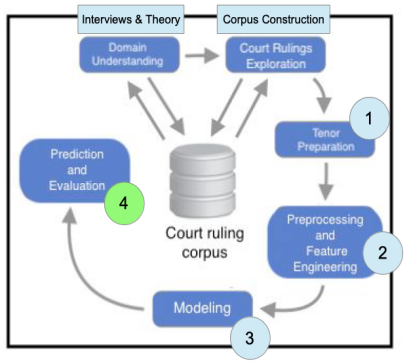
- Use the build feature vectors from the tenor text data and given instances of main, cost and enforceability decision as labels to build the classifiers

- Considered rule based approaches, multi class classification, multi label classification and tenor text generation with similarity check using given labels

- Went with the multi class classification approach with the following models
    1. Linear Support Vector
    2. Logistic Regression
    3. (Multinomial) Naive Bayes
    4. Random Forest

Curt Rouling Pipeline - Prediction

- Trained the classifiers on training data set



- Predicted legal consequences of unseen test data using the trained classifiers
- Made use of skicit learns build in classification report functionality for the split of the datasets
- Stratification to build very similiar distributions of labels across the training and test data sets
- Used cross validation with an folding factor of 10

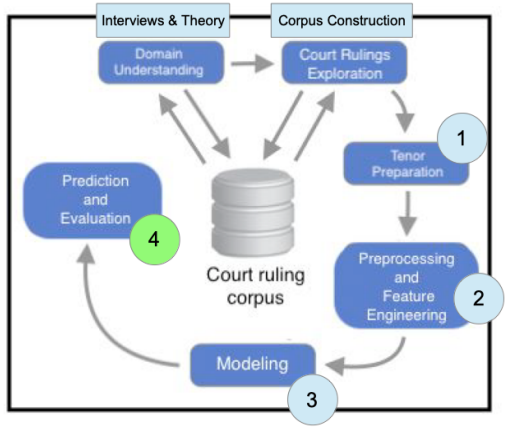
Curt Rouling Pipeline – Evaluation metrics

- Calculated the precision, recall, F1-score and accuracy of the predictions made
- Focused on the outcome of the F1-scores
- Also measured precision and recall

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$



- Calculated the accuracy scores, although it does not consider the tenor text data distribution
- Calculated also the macro and micro averages of the multi class classification for each tenor aspect

(If enough Time) Short Demo: Input, Algorithm and Outcome

Outline

Results – RQ1 – Extraction of legal consequences - (I)

Linear Support Vector Classifier

Results for Linear SVC

	precision	recall	f1-score	support
Allowance of the lawsuit	0.95	0.83	0.88	208
Dismissal of the lawsuit	0.92	0.97	0.94	325
Partial allowance	0.84	0.89	0.86	224
No main decision made	0.75	0.47	0.58	19
accuracy			0.90	776
macro avg	0.86	0.79	0.82	776
weighted avg	0.90	0.90	0.90	776

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.91	0.95	0.93	273
Defendant bearing the costs	0.90	0.91	0.91	212
Cost sharing	0.90	0.84	0.87	145
Cost cancellation	0.80	0.73	0.76	11
No cost decision made	0.96	0.95	0.96	135
accuracy			0.91	776
macro avg	0.89	0.87	0.88	776
weighted avg	0.91	0.91	0.91	776

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.93	0.90	0.92	133
PE against security deposit	0.98	0.96	0.97	123
Provisionally Avertable	0.98	0.99	0.99	279
No enforceability decision made	0.96	0.97	0.96	241
accuracy			0.96	776
macro avg	0.96	0.96	0.96	776
weighted avg	0.96	0.96	0.96	776

Micro averages for Random Forest Classifier – main, cost and enforceability decision

Precision	–	Recall	–	F1 Score
(0.896907216494845	4,	0.896907216494845	4,	0.8969072164948453, None)
(0.91365979381443	3,	0.91365979381443	3,	0.913659793814433, None)
(0.963917525773195	9,	0.963917525773195	9,	0.9639175257731959, None)

Logistic Regression Classifier

Results for Logistic Regression

	precision	recall	f1-score	support
Allowance of the lawsuit	0.93	0.82	0.87	208
Dismissal of the lawsuit	0.88	0.99	0.93	325
Partial allowance	0.85	0.84	0.85	224
No main decision made	1.00	0.26	0.42	19
accuracy			0.88	776
macro avg	0.91	0.73	0.77	776
weighted avg	0.89	0.88	0.88	776

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.89	0.95	0.91	273
Defendant bearing the costs	0.85	0.92	0.88	212
Cost sharing	0.87	0.77	0.82	145
Cost cancellation	1.00	0.09	0.17	11
No cost decision made	0.99	0.91	0.95	135
accuracy			0.89	776
macro avg	0.92	0.73	0.75	776
weighted avg	0.89	0.89	0.88	776

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.93	0.86	0.90	133
PE against security deposit	0.97	0.95	0.96	123
Provisionally Avertable	0.98	1.00	0.99	279
No enforceability decision made	0.94	0.97	0.96	241
accuracy			0.96	776
macro avg	0.96	0.95	0.95	776
weighted avg	0.96	0.96	0.96	776

Micro averages for Random Forest Classifier – main, cost and enforceability decision

Precision	–	Recall	–	F1 Score
(0.882731958762886	6,	0.882731958762886	6,	0.8827319587628866, None)
(0.889175257731958	7,	0.889175257731958	7,	0.8891752577319587, None)
(0.960051546391752	6,	0.960051546391752	6,	0.9600515463917526, None)

Results – RQ1 – Extraction of legal consequences (II)

Multinomial Naive Bayes Classifier

Results for Multinomial Naive Bayes Classifier

	precision	recall	f1-score	support
Allowance of the lawsuit	0.80	0.59	0.68	208
Dismissal of the lawsuit	0.87	0.96	0.91	325
Partial allowance	0.66	0.77	0.71	224
No main decision made	1.00	0.26	0.42	19
accuracy			0.79	776
macro avg	0.83	0.65	0.68	776
weighted avg	0.80	0.79	0.78	776

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.80	0.90	0.85	273
Defendant bearing the costs	0.81	0.73	0.77	212
Cost sharing	0.79	0.72	0.75	145
Cost cancellation	0.00	0.00	0.00	11
No cost decision made	0.81	0.91	0.86	135
accuracy			0.81	776
macro avg	0.65	0.65	0.65	776
weighted avg	0.80	0.81	0.80	776

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.81	0.64	0.71	133
PE against security deposit	0.77	0.93	0.84	123
Provisionally Avertable	0.93	0.97	0.95	279
No enforceability decision made	0.96	0.91	0.93	241
accuracy			0.89	776
macro avg	0.87	0.86	0.86	776
weighted avg	0.89	0.89	0.89	776

Micro averages for Multinomial Naive Bayes Classifier – main, cost and enforceability decision

Precision	–	Recall	–	F1 Score
(0.78865979381443,		0.78865979381443,		0.78865979381443, None)
(0.8067010309278351,		0.8067010309278351,		0.8067010309278352, None)
(0.8891752577319587,		0.8891752577319587,		0.8891752577319587, None)

Random Forest Classifier

Results for Logistic Regression

	precision	recall	f1-score	support
Allowance of the lawsuit	0.93	0.82	0.87	208
Dismissal of the lawsuit	0.88	0.99	0.93	325
Partial allowance	0.85	0.84	0.85	224
No main decision made	1.00	0.26	0.42	19
accuracy			0.88	776
macro avg	0.91	0.73	0.77	776
weighted avg	0.89	0.88	0.88	776

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.89	0.95	0.91	273
Defendant bearing the costs	0.85	0.92	0.88	212
Cost sharing	0.87	0.77	0.82	145
Cost cancellation	1.00	0.09	0.17	11
No cost decision made	0.99	0.91	0.95	135
accuracy			0.89	776
macro avg	0.92	0.73	0.75	776
weighted avg	0.89	0.89	0.88	776

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.93	0.86	0.90	133
PE against security deposit	0.97	0.95	0.96	123
Provisionally Avertable	0.98	1.00	0.99	279
No enforceability decision made	0.94	0.97	0.96	241
accuracy			0.96	776
macro avg	0.96	0.95	0.95	776
weighted avg	0.96	0.96	0.96	776

Micro averages for Random Forest Classifier – main, cost and enforceability decision

Precision	–	Recall	–	F1 Score
(0.8878865979381443,		0.8878865979381443,		0.8878865979381443, None)
(0.8904639175257731,		0.8904639175257731,		0.8904639175257731, None)
(0.9536082474226805,		0.9536082474226805,		0.9536082474226805, None)

Results – RQ2 – Derivation of legal consequences (LC) (I)

Linear Support Vector Classifier from the 3rd Instance tenor text. . .

Derivation of the 2nd Instance LC

Results for Linear SVC predicting the legal consequences from the appeal instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.10	0.33	0.15	9
Dismissal of the lawsuit	0.60	0.42	0.49	36
Partial allowance	0.50	0.30	0.37	27
No main decision made	0.00	0.00	0.00	1
accuracy			0.36	73
macro avg	0.30	0.26	0.25	73
weighted avg	0.49	0.36	0.40	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.58	0.22	0.32	32
Defendant bearing the costs	0.50	0.43	0.46	21
Cost sharing	1.00	0.06	0.12	16
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.05	0.67	0.09	3
accuracy			0.26	73
macro avg	0.43	0.28	0.20	73
weighted avg	0.62	0.26	0.30	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	4
PE against security deposit	0.00	0.00	0.00	2
Provisionally Avertable	0.00	0.00	0.00	65
No enforceability decision made	0.03	1.00	0.05	2
accuracy			0.03	73
macro avg	0.01	0.25	0.01	73
weighted avg	0.00	0.03	0.00	73

Derivation of the 1st Instance LC

Results for Linear SVC predicting the legal consequences from the first instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.24	0.47	0.31	17
Dismissal of the lawsuit	0.50	0.38	0.43	32
Partial allowance	0.27	0.19	0.22	21
No main decision made	0.00	0.00	0.00	3
accuracy			0.33	73
macro avg	0.25	0.26	0.24	73
weighted avg	0.35	0.33	0.32	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.50	0.25	0.33	32
Defendant bearing the costs	0.60	0.41	0.49	22
Cost sharing	0.00	0.00	0.00	15
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.08	1.00	0.15	3
accuracy			0.27	73
macro avg	0.24	0.33	0.19	73
weighted avg	0.40	0.27	0.30	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	2
PE against security deposit	0.00	0.00	0.00	48
Provisionally Avertable	0.00	0.00	0.00	19
No enforceability decision made	0.05	1.00	0.10	4
accuracy			0.05	73
macro avg	0.01	0.25	0.03	73
weighted avg	0.00	0.05	0.01	73

Results – RQ2 – Derivation of legal consequences (LC) (II)

Logistic Regression Classifier from the 3rd Instance tenor text. . .

Derivation of the 2nd Instance LC

Results for Logistic Regression Classifier predicting the legal consequences from the appeal instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.11	0.44	0.18	9
Dismissal of the lawsuit	0.61	0.39	0.47	36
Partial allowance	0.57	0.30	0.39	27
No main decision made	0.00	0.00	0.00	1
accuracy			0.36	73
macro avg	0.32	0.28	0.26	73
weighted avg	0.53	0.36	0.40	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.75	0.19	0.30	32
Defendant bearing the costs	0.53	0.48	0.50	21
Cost sharing	1.00	0.06	0.12	16
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.07	1.00	0.12	3
accuracy			0.27	73
macro avg	0.47	0.35	0.21	73
weighted avg	0.70	0.27	0.31	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	4
PE against security deposit	0.00	0.00	0.00	2
Provisionally Avertable	0.00	0.00	0.00	65
No enforceability decision made	0.03	1.00	0.05	2
accuracy			0.03	73
macro avg	0.01	0.25	0.01	73
weighted avg	0.00	0.03	0.00	73

Derivation of the 1st Instance LC

Results for Logistic Regression Classifier predicting the legal consequences from the first instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.25	0.53	0.34	17
Dismissal of the lawsuit	0.50	0.34	0.41	32
Partial allowance	0.20	0.14	0.17	21
No main decision made	0.00	0.00	0.00	3
accuracy			0.32	73
macro avg	0.24	0.25	0.23	73
weighted avg	0.33	0.32	0.31	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.50	0.22	0.30	32
Defendant bearing the costs	0.56	0.41	0.47	22
Cost sharing	0.00	0.00	0.00	15
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.07	1.00	0.13	3
accuracy			0.26	73
macro avg	0.23	0.33	0.18	73
weighted avg	0.39	0.26	0.28	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	2
PE against security deposit	0.00	0.00	0.00	48
Provisionally Avertable	0.00	0.00	0.00	19
No enforceability decision made	0.05	1.00	0.10	4
accuracy			0.05	73
macro avg	0.01	0.25	0.03	73
weighted avg	0.00	0.05	0.01	73

Results – RQ2 – Derivation of legal consequences (LC) (III)

Multinomial Naive Bayes Classifier from the 3rd Instance tenor text. . .

Derivation of the 2nd Instance LC

Results for Multinomial Naive Bayes Classifier predicting the legal consequences from the appeal instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.14	0.44	0.21	9
Dismissal of the lawsuit	0.56	0.39	0.46	36
Partial allowance	0.58	0.41	0.48	27
No main decision made	0.00	0.00	0.00	1
accuracy			0.40	73
macro avg	0.32	0.31	0.29	73
weighted avg	0.51	0.40	0.43	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.67	0.19	0.29	32
Defendant bearing the costs	0.47	0.43	0.45	21
Cost sharing	1.00	0.12	0.22	16
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.07	1.00	0.13	3
accuracy			0.27	73
macro avg	0.44	0.35	0.22	73
weighted avg	0.65	0.27	0.31	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	4
PE against security deposit	0.00	0.00	0.00	2
Provisionally Avertable	0.00	0.00	0.00	65
No enforceability decision made	0.03	1.00	0.05	2
accuracy			0.03	73
macro avg	0.01	0.25	0.01	73
weighted avg	0.00	0.03	0.00	73

Derivation of the 1st Instance LC

Results for Multinomial Naive Bayes Classifier predicting the legal consequences from the first instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.24	0.47	0.32	17
Dismissal of the lawsuit	0.48	0.34	0.40	32
Partial allowance	0.18	0.14	0.16	21
No main decision made	0.00	0.00	0.00	3
accuracy			0.30	73
macro avg	0.22	0.24	0.22	73
weighted avg	0.32	0.30	0.30	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.38	0.16	0.22	32
Defendant bearing the costs	0.50	0.41	0.45	22
Cost sharing	0.00	0.00	0.00	15
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.05	0.67	0.09	3
accuracy			0.22	73
macro avg	0.19	0.25	0.15	73
weighted avg	0.32	0.22	0.24	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	2
PE against security deposit	0.00	0.00	0.00	48
Provisionally Avertable	0.00	0.00	0.00	19
No enforceability decision made	0.05	1.00	0.10	4
accuracy			0.05	73
macro avg	0.01	0.25	0.03	73
weighted avg	0.00	0.05	0.01	73

Results – RQ2 – Derivation of legal consequences (LC) (IV)

Random Forest Classifier from the 3rd Instance tenor text. . .

Derivation of the 2nd Instance LC

Results for Random Forest Classifier predicting the legal consequences from the appeal instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.12	0.44	0.20	9
Dismissal of the lawsuit	0.54	0.36	0.43	36
Partial allowance	0.53	0.30	0.38	27
No main decision made	0.00	0.00	0.00	1
accuracy			0.34	73
macro avg	0.30	0.28	0.25	73
weighted avg	0.48	0.34	0.38	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.64	0.22	0.33	32
Defendant bearing the costs	0.50	0.43	0.46	21
Cost sharing	1.00	0.06	0.12	16
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.05	0.67	0.09	3
accuracy			0.26	73
macro avg	0.44	0.28	0.20	73
weighted avg	0.64	0.26	0.30	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	4
PE against security deposit	0.00	0.00	0.00	2
Provisionally Avertable	0.00	0.00	0.00	65
No decision made	0.03	1.00	0.05	2
accuracy			0.03	73
macro avg	0.01	0.25	0.01	73
weighted avg	0.00	0.03	0.00	73

Derivation of the 1st Instance LC

Results for Random Forest Classifier predicting the legal consequences from the first instance court ruling

	precision	recall	f1-score	support
Allowance of the lawsuit	0.21	0.41	0.27	17
Dismissal of the lawsuit	0.48	0.38	0.42	32
Partial allowance	0.21	0.14	0.17	21
No main decision made	0.00	0.00	0.00	3
accuracy			0.30	73
macro avg	0.23	0.23	0.22	73
weighted avg	0.32	0.30	0.30	73

	precision	recall	f1-score	support
Plaintiff bearing the costs	0.50	0.25	0.33	32
Defendant bearing the costs	0.56	0.41	0.47	22
Cost sharing	0.00	0.00	0.00	15
Cost cancellation	0.00	0.00	0.00	1
No cost decision made	0.08	1.00	0.15	3
accuracy			0.27	73
macro avg	0.23	0.33	0.19	73
weighted avg	0.39	0.27	0.29	73

	precision	recall	f1-score	support
Provisional Enforceability (PE)	0.00	0.00	0.00	2
PE against security deposit	0.00	0.00	0.00	48
Provisionally Avertable	0.00	0.00	0.00	19
No enforceability decision made	0.05	1.00	0.10	4
accuracy			0.05	73
macro avg	0.01	0.25	0.03	73
weighted avg	0.00	0.05	0.01	73

Results

- Need for reduction of manual labor and supportive tool verified through judges interviews
- Extraction of legal consequences from the tenor text through multi class classification works well
- Derivation of legal consequences from the revision instance performs bad using only tenor text data

Discussion

- More intelligent searching and filtering functionalities would be helpful for legal domain experts, e.g. Juris.de

- Considering other parts of a court ruling will be needed for the derivation of legal consequences

- More advanced machine learning approaches like LSTM, Transformer or Few Shot Learning could be benefitial for the extraction of legal consequences from tenor text

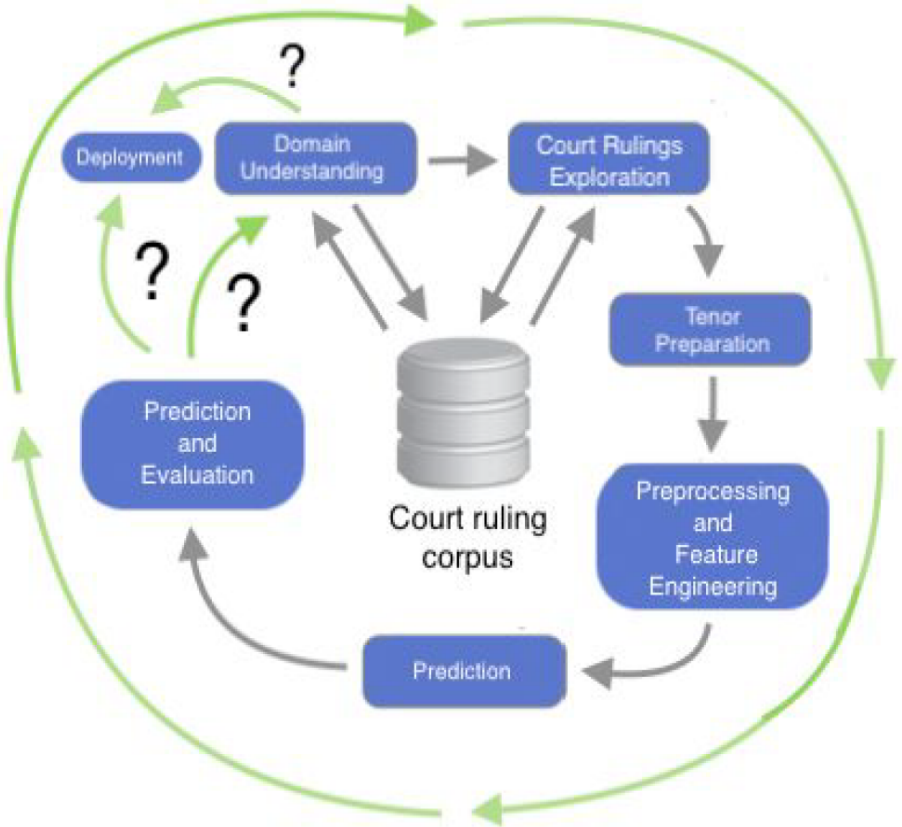
Outline

Conclusion

- Established baseline classifiers work well for the extraction of legal consequences (RQ1)
- Classifiers performed not good for the derivation of the legal consequences (RQ2)
- Further consideration of other court ruling parts needed for the derivation of legal consequences
- Interview results showed that the tenor is not often needed in practice by judges

Future Work

- Extraction of legal consequences could be done by text generation and text similarity checks
- Using BERT models specific to legal domain like LegalBERT for text generation with One Shot Learning (Siamese Network) for text similarity checks
- Court ruling pipeline could be extended to deliver new insights back to the domain continuously



Thank you for your attention !

Suggestions ?

Questions ?

Remarks/Feedback/Opinions ?

Outline

References (I)

[Snorkel.ai]
https://snorkel.ai/ and https://github.com/snorkel-team/snorkel started in Stanford 2016

[Erium], https://erium.de/ combination of small data sets and domain expert knowledge

[HyPerDB], https://hyper-db.de/ main-memory-based relational DBMS for mixed OLTP and OLAP workloads

[BmJV1] Bundesministeriumr für Justiz undVerbraucherschutz,
https://www.bmjv.de/SharedDocs/Archiv/Downloads/Schaubild_Gerichtsaufbau_Deutsch.pdf?__blob=publicationFile&v=3

[ZPO] §313, ZPO; https://www.gesetze-im-internet.de/zpo/__313a.html

[Hofmann] Repititorium Hofmann, https://www.repetitorium-hofmann.de/pdf/einfuehrung_ins_referendariat_zpo_skript.pdf

[Kurpat] Ralf Kurpat, Peter Siegburg, Das Zivilurteil, page 23 – 83. https://www.beck-elibrary.de/10.15358/9783800658411-23/b-die-urteilsformel

[BmJV1] Bundesministeriumr für Justiz undVerbraucherschutz,
https://www.bmjv.de/SharedDocs/Downloads/DE/PDF/Anzahl_der_Gerichte_des_Bundes_und_der_Laender.pdf?__blob=publicationFile&v=2

References (II)

[BmJV2] Richterstatistik 2018. (PDF) Bundesamt für Justiz, 31. Dezember 2018, abgerufen am 8. September 2020. https://www.bundesjustizamt.de/DE/SharedDocs/Publikationen/Justizstatistik/Richterstatistik_2018.pdf?__blob=publicationFile&v=3

[La19a] User Study on Selection Search and Semantic Text Matching in German Tenancy Law, 2019

[Lynx] https://lynx-project.eu/project/consortium

[Leitner] A Dataset of German Legal Documents for Named Entity Recognition, Elena Leitner, Georg Rehm, Julian Moreno-Schneider

[Urchs et. al] "Towards Classifying Parts of German Legal Writing Styles in German Legal Judgments." 2020 10th International Conference on Advanced Computer Information Technologies (ACIT) . IEEE, 2020.

Mario Turić

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289.
Fax +49.89.289.17136

wwwmatthes.in.tum.de