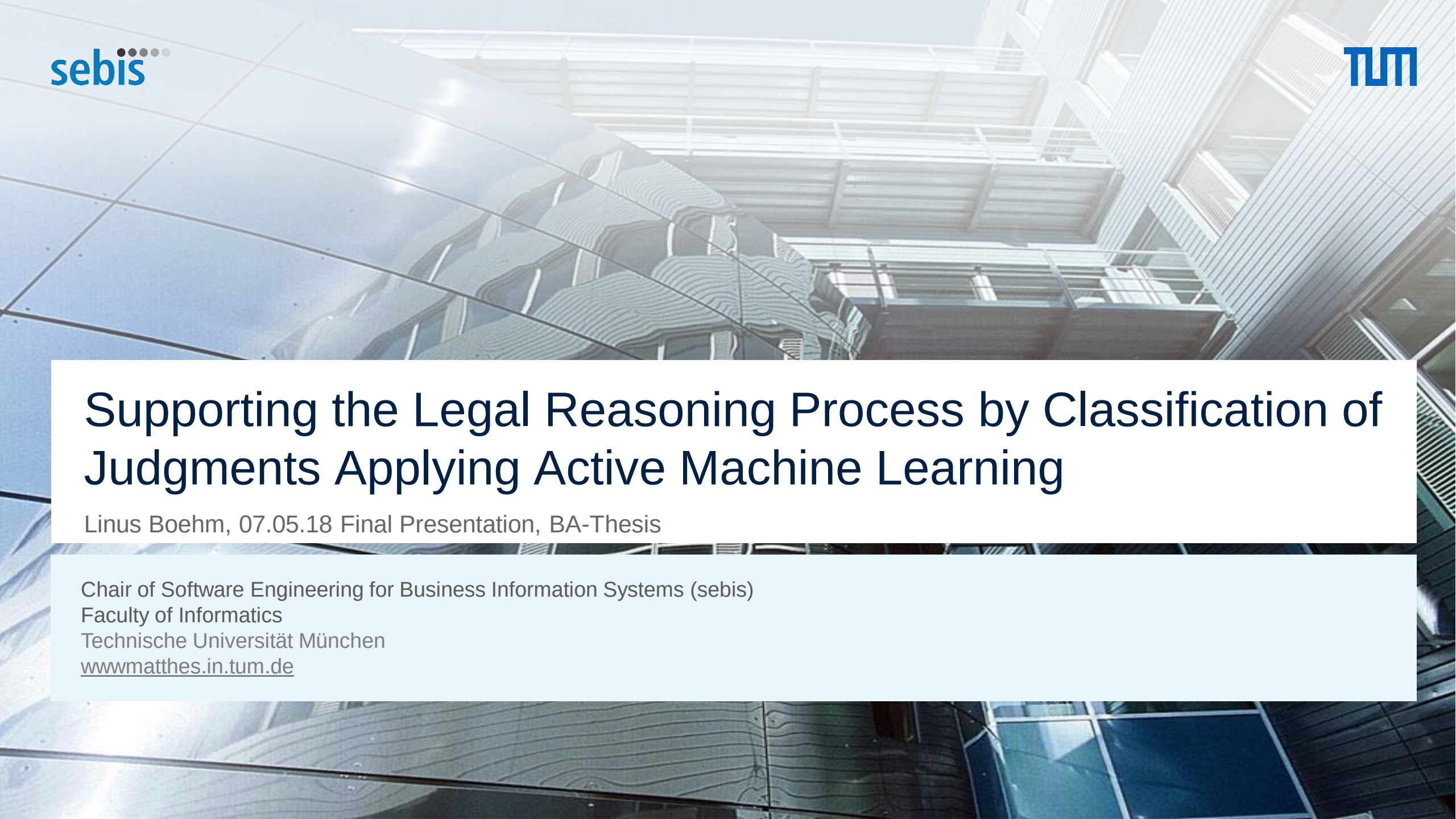


Supporting the Legal Reasoning Process by Classification of Judgments Applying Active Machine Learning

Linus Boehm, 07.05.18 Final Presentation, BA-Thesis

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

Outline

Motivation

Legal Documents

- are growing and changing: *2013-2017 period more than 550 adopted laws [1]*
- getting more complex

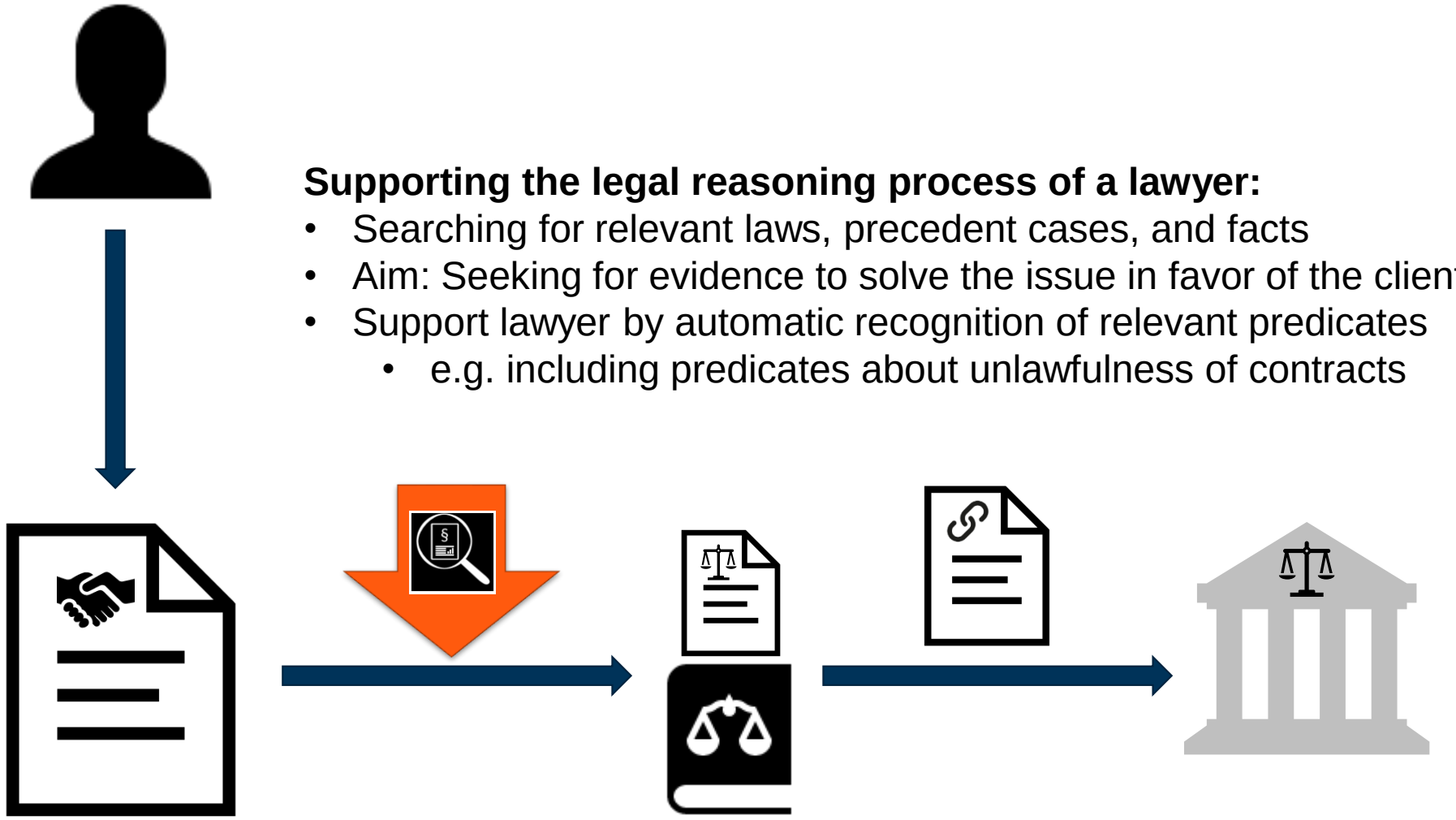
Document Discovery

- often done manually (knowledge-intensive)
- very time-consuming and costly [2]

Supporting the legal reasoning process by classification of Judgments

- common law relies heavily on precedents
- judgments relevant in (German) civil law?
  - further development of the law by judges
  - Case law is often applied in dynamic areas of life
    - German landlord and tenant law is largely shaped by case law

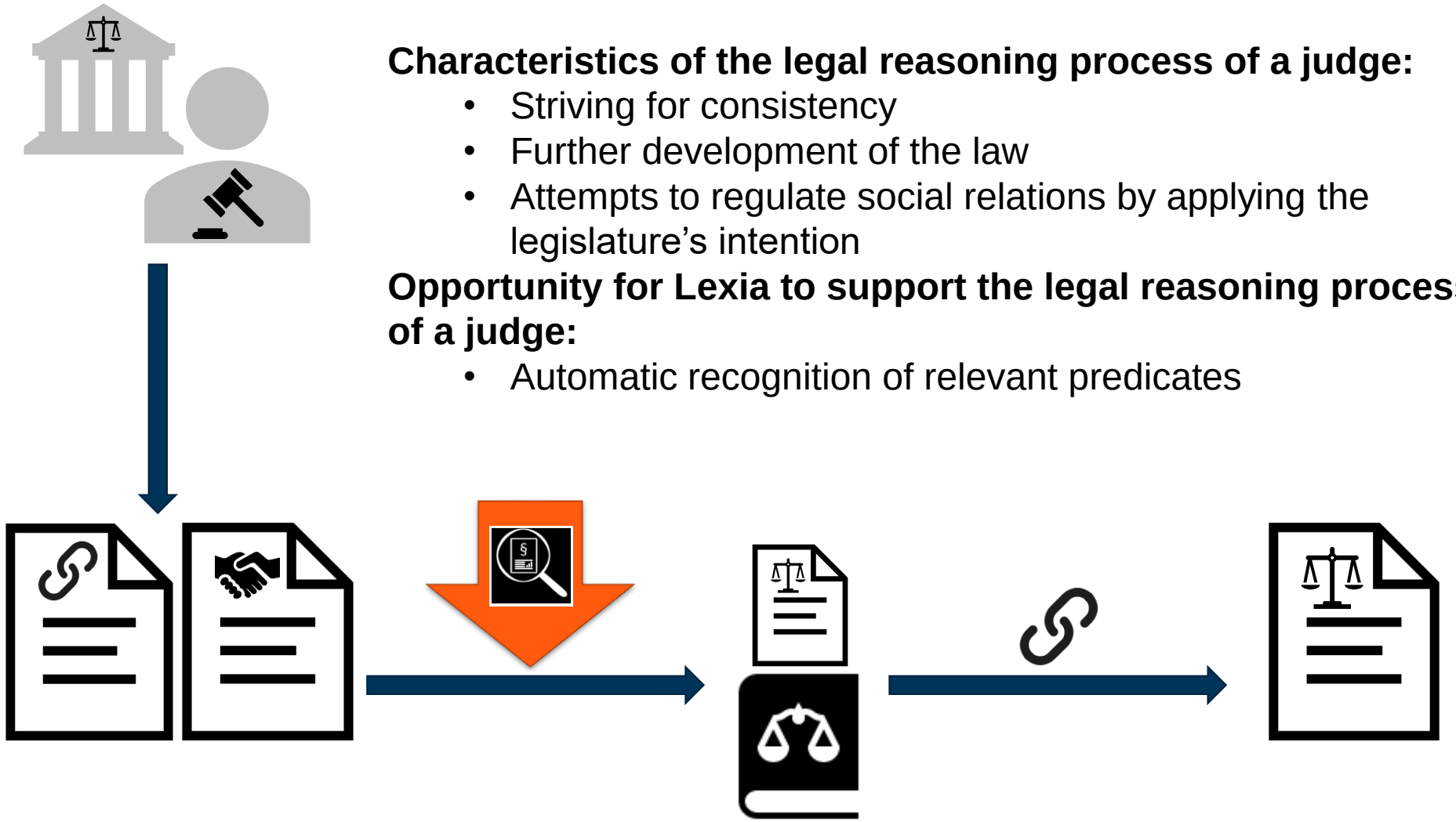
Legal Reasoning Process – Target



Supporting the legal reasoning process of a lawyer:

- Searching for relevant laws, precedent cases, and facts
- Aim: Seeking for evidence to solve the issue in favor of the client
- Support lawyer by automatic recognition of relevant predicates
  - e.g. including predicates about unlawfulness of contracts

Legal Reasoning Process – Target



Characteristics of the legal reasoning process of a judge:

- Striving for consistency
- Further development of the law
- Attempts to regulate social relations by applying the legislature's intention

Opportunity for Lexia to support the legal reasoning process of a judge:

- Automatic recognition of relevant predicates

Approach

Manuel Annotation of Judgments

Sentences about the unlawfulness of contract clauses

Implementation of a BTC for judgments

Adaption of Lexia and LexML

Evaluation

Evaluating common performance indicators like f1score, ROC, recall and precision

Manual Annotation Process

Contract clause	+	legal justification	+	unlawfullness
e.g: Vertragsklausel Formularklausel §x des Vertrages		e.g: Paragraph of Civil Code Precendant of the BGH		e.g: Unwirksam Nichtig Nichtigkeit

Example:

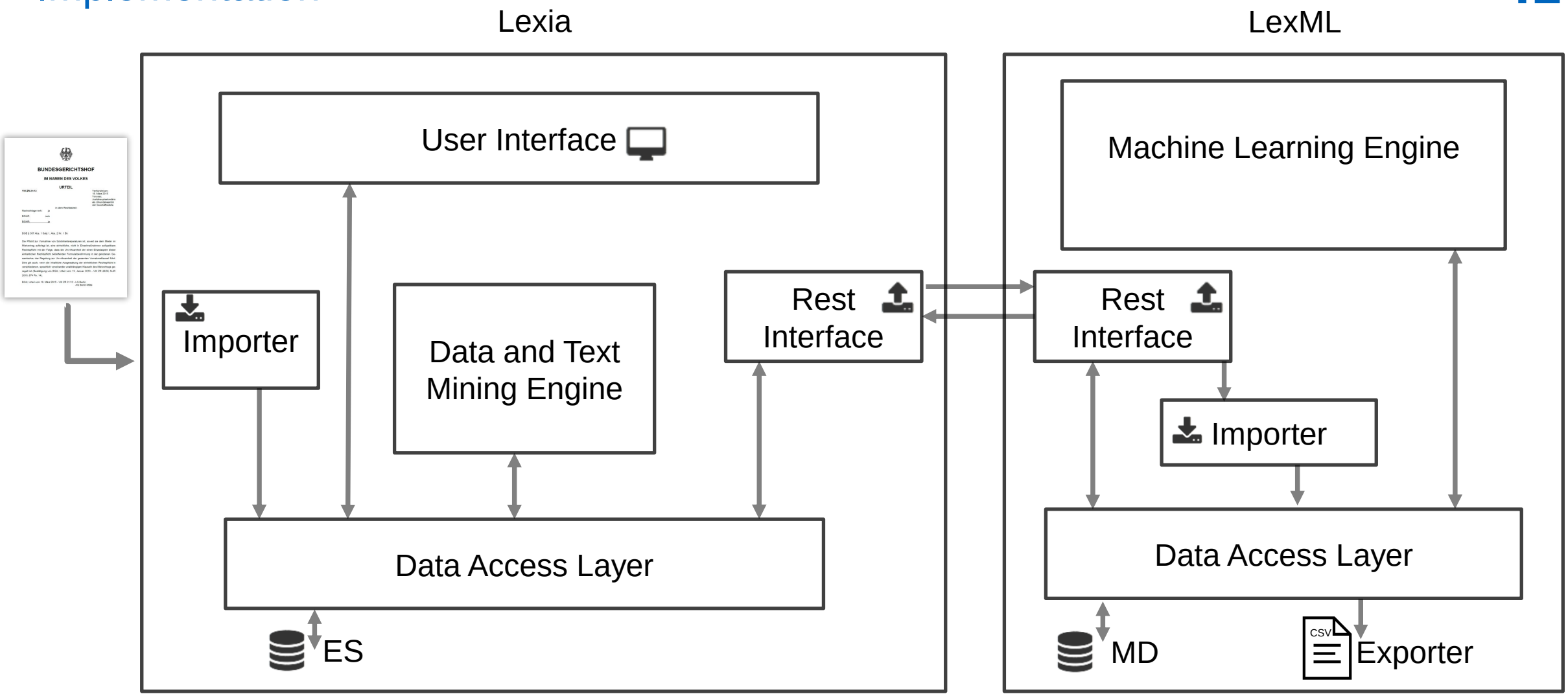
[…] entsprechende Formularklausel wegen unangemessener Benachteiligung des Mieters nach § 307 BGB unwirksam […]

Problems with the Annotation „Rule“:

- Sometimes one Part is mentioned in following or previous sentence -> partly causing the False Positives

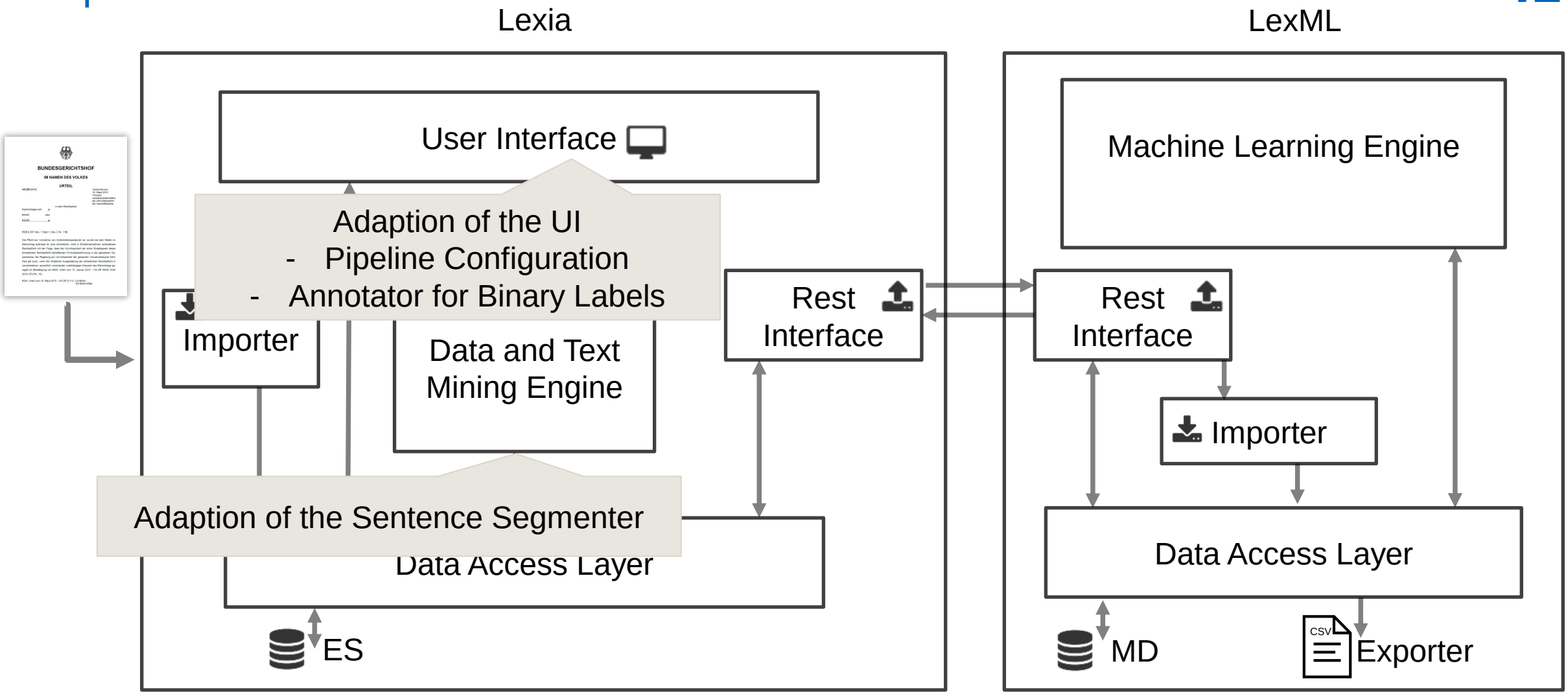
Implementation

Implementation

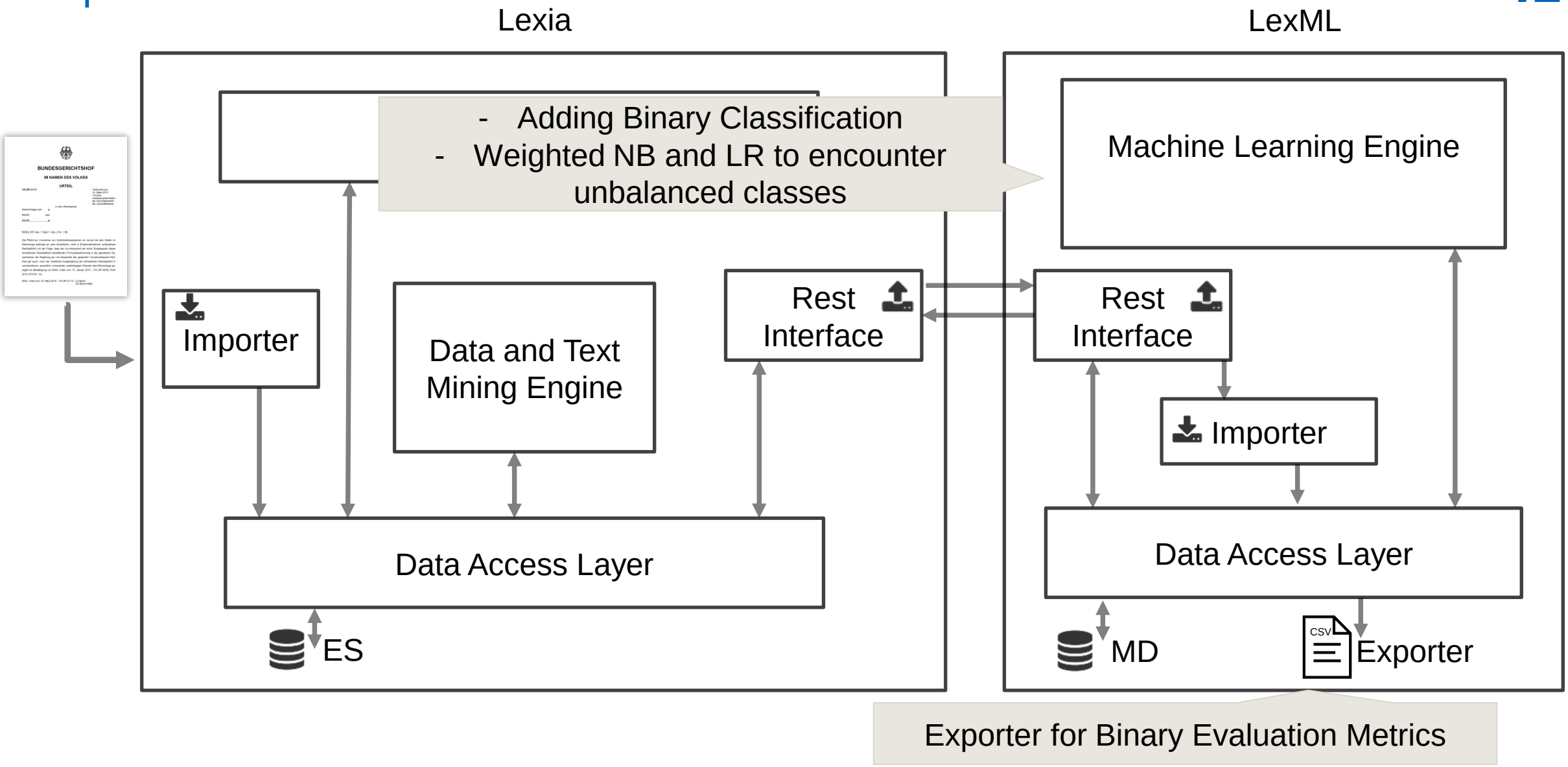


Own illustration based on Muhr [3]

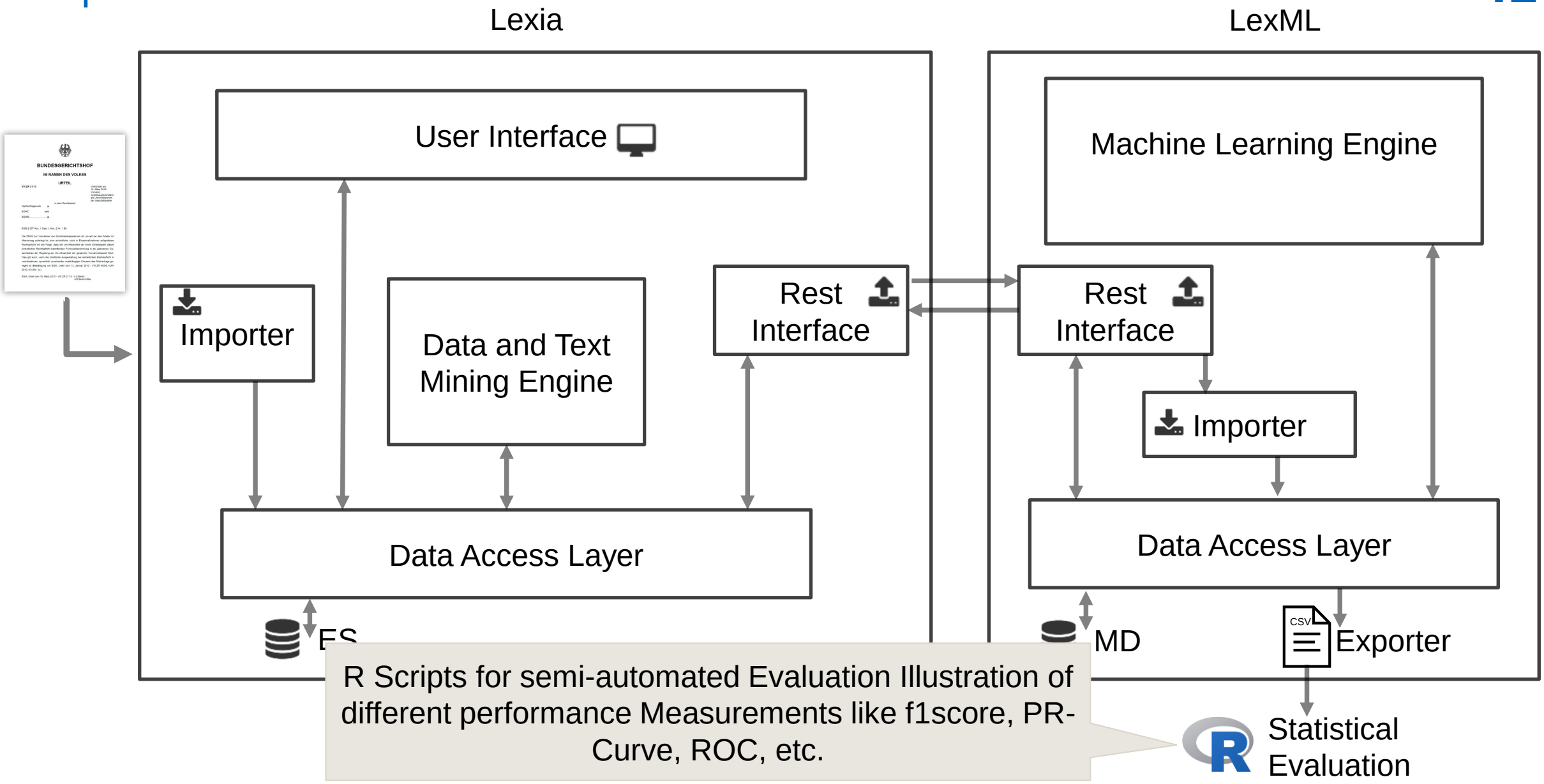
Implementation



Implementation



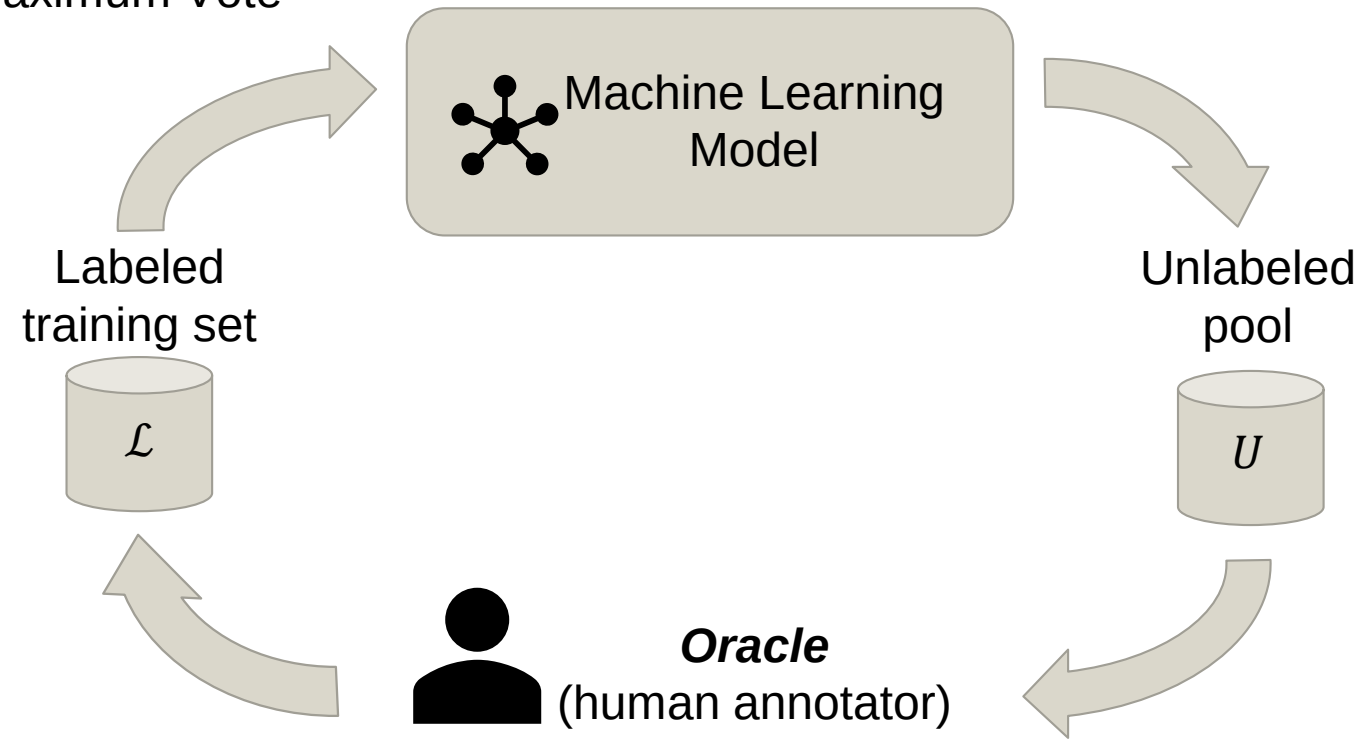
Implementation



Active Learning – Setting

AML Pipline Stages:

- Preprocessing: Tokenizer, Stopwords Removing
- Vector Representation: Binary Word Representation
- Classifier: Multinominal Naive Bayes (Weighted)
- Query Strategy: Uncertainty Sampling (Maximum Vote Entropy)

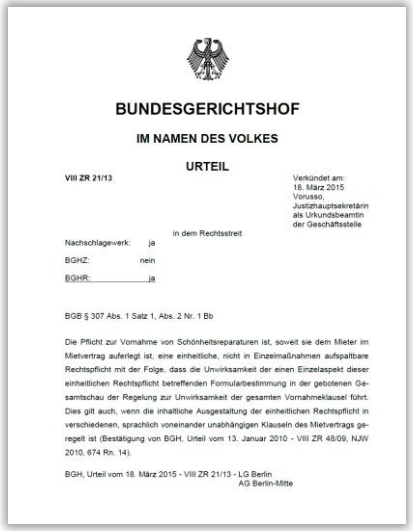


Data Set

Data Set

Raw Documents

- Over 800 imported and annotated judgments
- BGH judgments of the VIII. Civil Senate
  - specialized in law on the sale of goods, landlord and tenancy law



BUNDESGERICHTSHOF

IM NAMEN DES VOLKES

URTEIL

VIII ZR 21/13

Verkündet am:
18. März 2015
Vorusso,
Justizhauptsekretärin
als Urkundsbeamtin
der Geschäftsstelle

in dem Rechtsstreit

Nachschlagewerk: ja
BGHZ: nein
BGHR: ja

BGB § 307 Abs. 1 Satz 1, Abs. 2 Nr. 1 Bb

Die Pflicht zur Vornahme von Schönheitsreparaturen ist, soweit sie dem Mieter im Mietvertrag auferlegt ist, eine einheitliche, nicht in Einzelmaßnahmen aufspaltbare Rechtspflicht mit der Folge, dass die Unwirksamkeit der einen Einzelaspekt dieser einheitlichen Rechtspflicht betreffenden Formularbestimmung in der gebotenen Gesamtschau der Regelung zur Unwirksamkeit der gesamten Vornahmeklausel führt. Dies gilt auch, wenn die inhaltliche Ausgestaltung der einheitlichen Rechtspflicht in verschiedenen, sprachlich voneinander unabhängigen Klauseln des Mietvertrags geregelt ist (Bestätigung von BGH, Urteil vom 13. Januar 2010 - VIII ZR 48/09, NJW 2010, 674 Rn. 14).

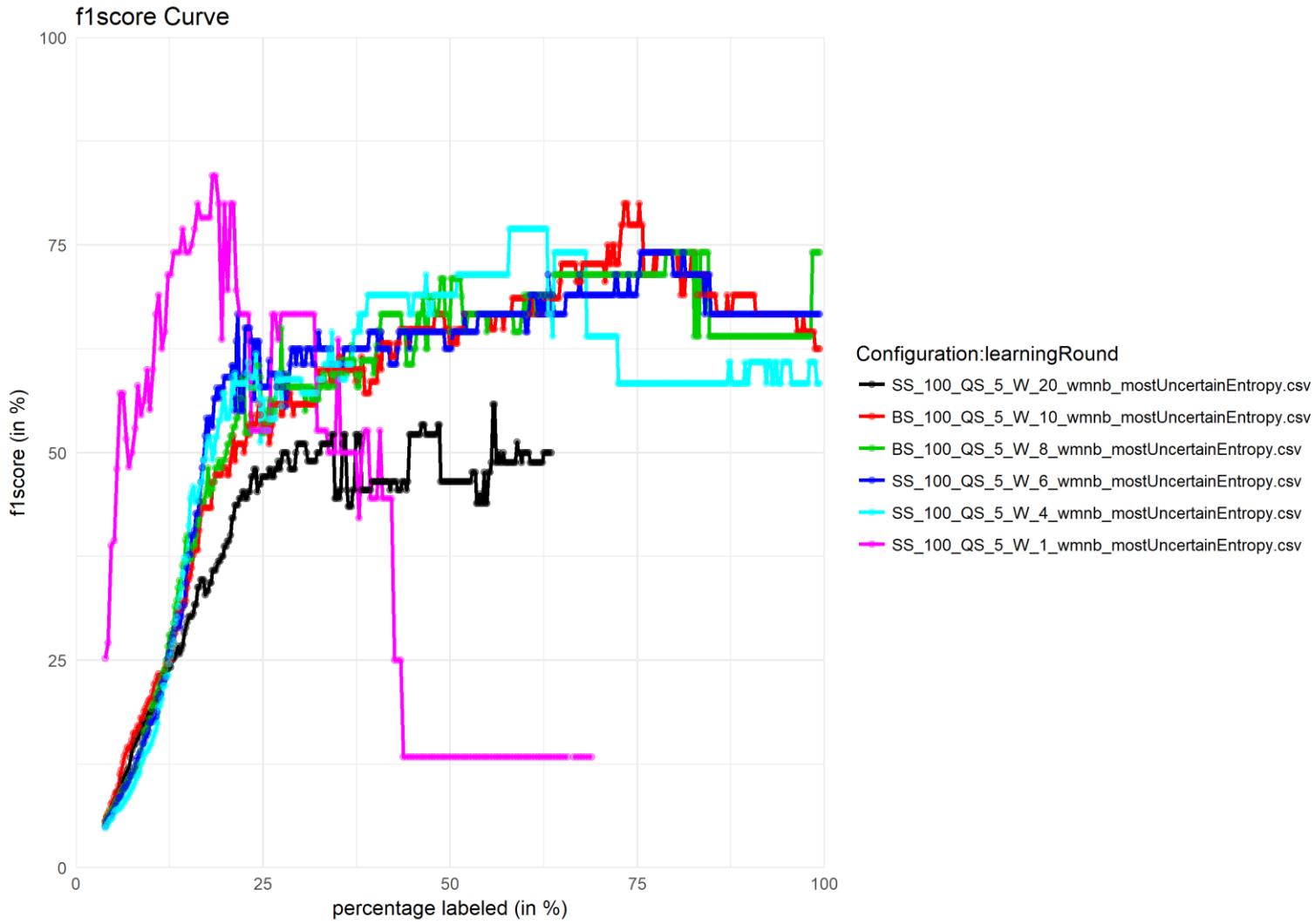
BGH, Urteil vom 18. März 2015 - VIII ZR 21/13 - LG Berlin
AG Berlin-Mitte

Training & Test Set

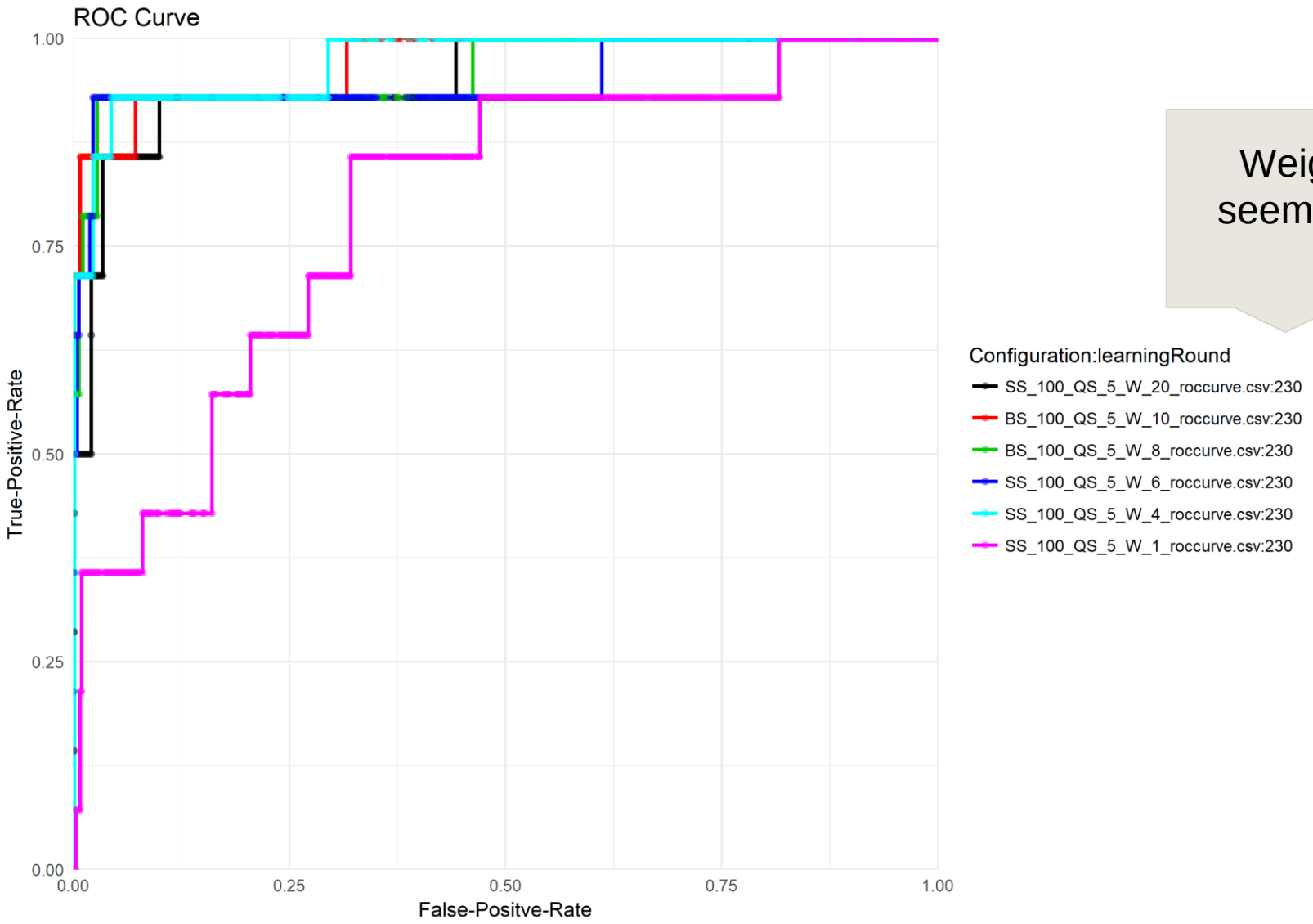
- Using a subset of 82 documents
- Only Tenor and Reasoning part imported
- 3135 Sentences of which are 72 True (2,3%)
- 80/20 partition

Evaluation

Discover weight factor

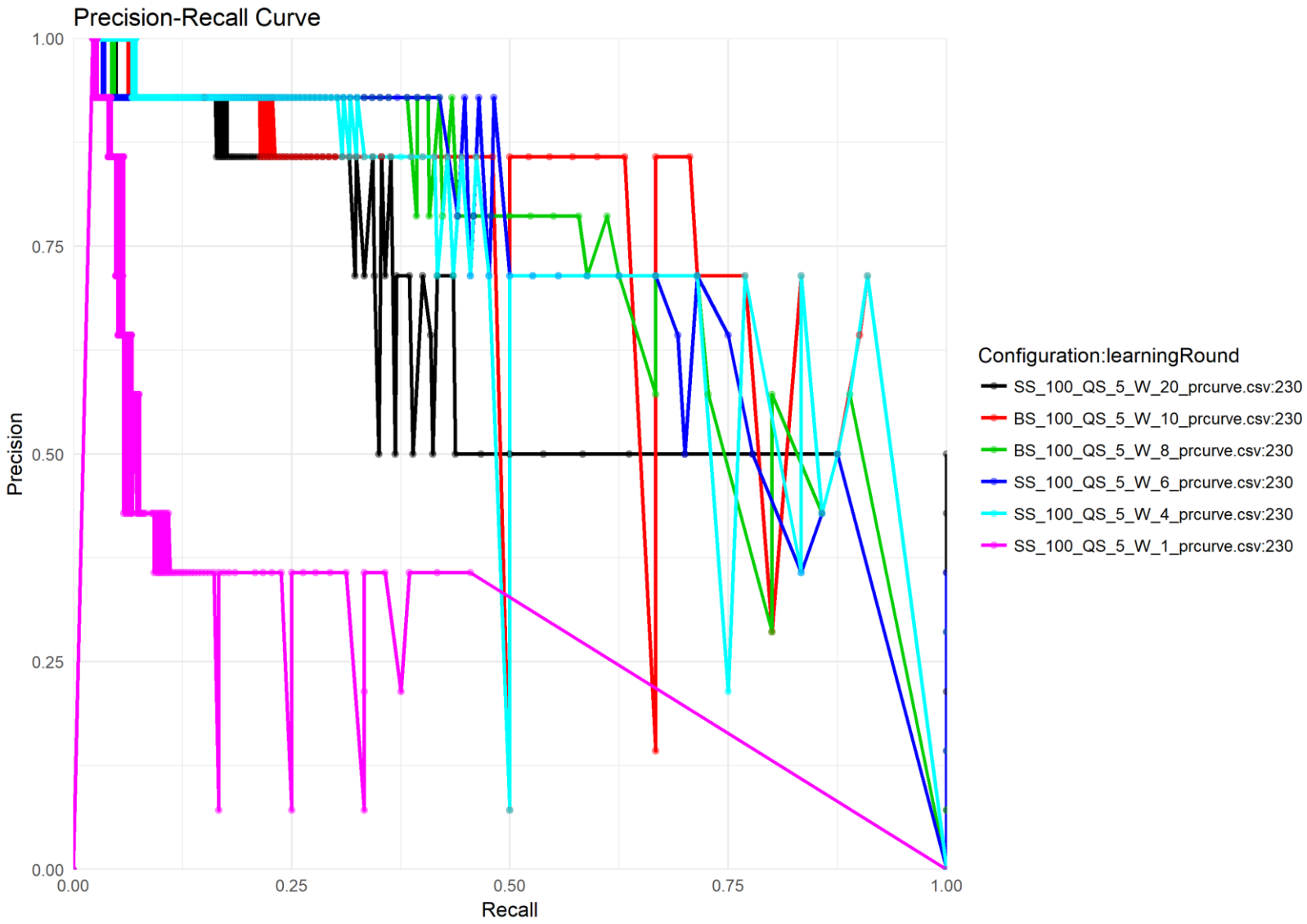


Discover weight factor



Weight 20/0.05 seems to perform well

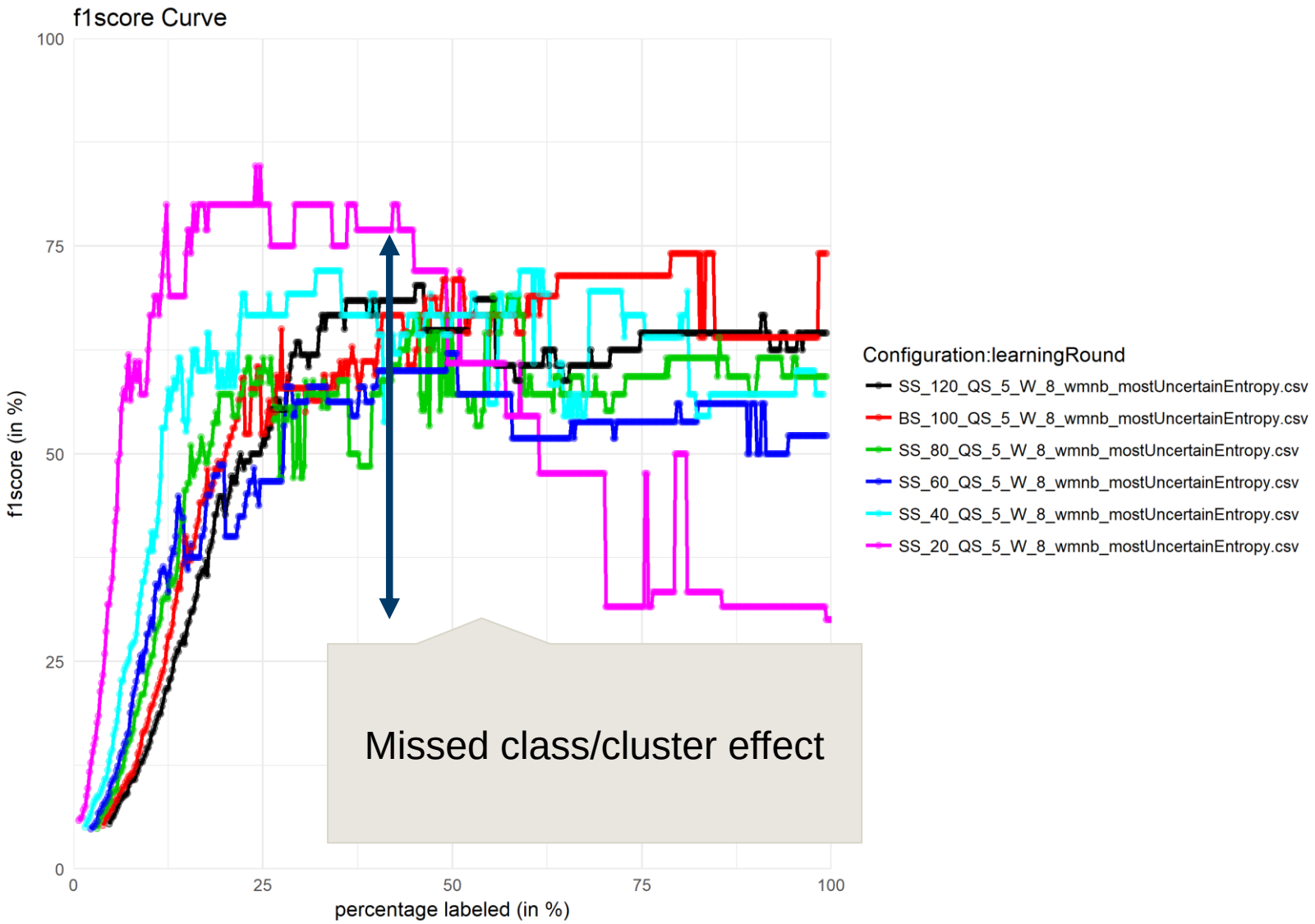
Discover weight factor



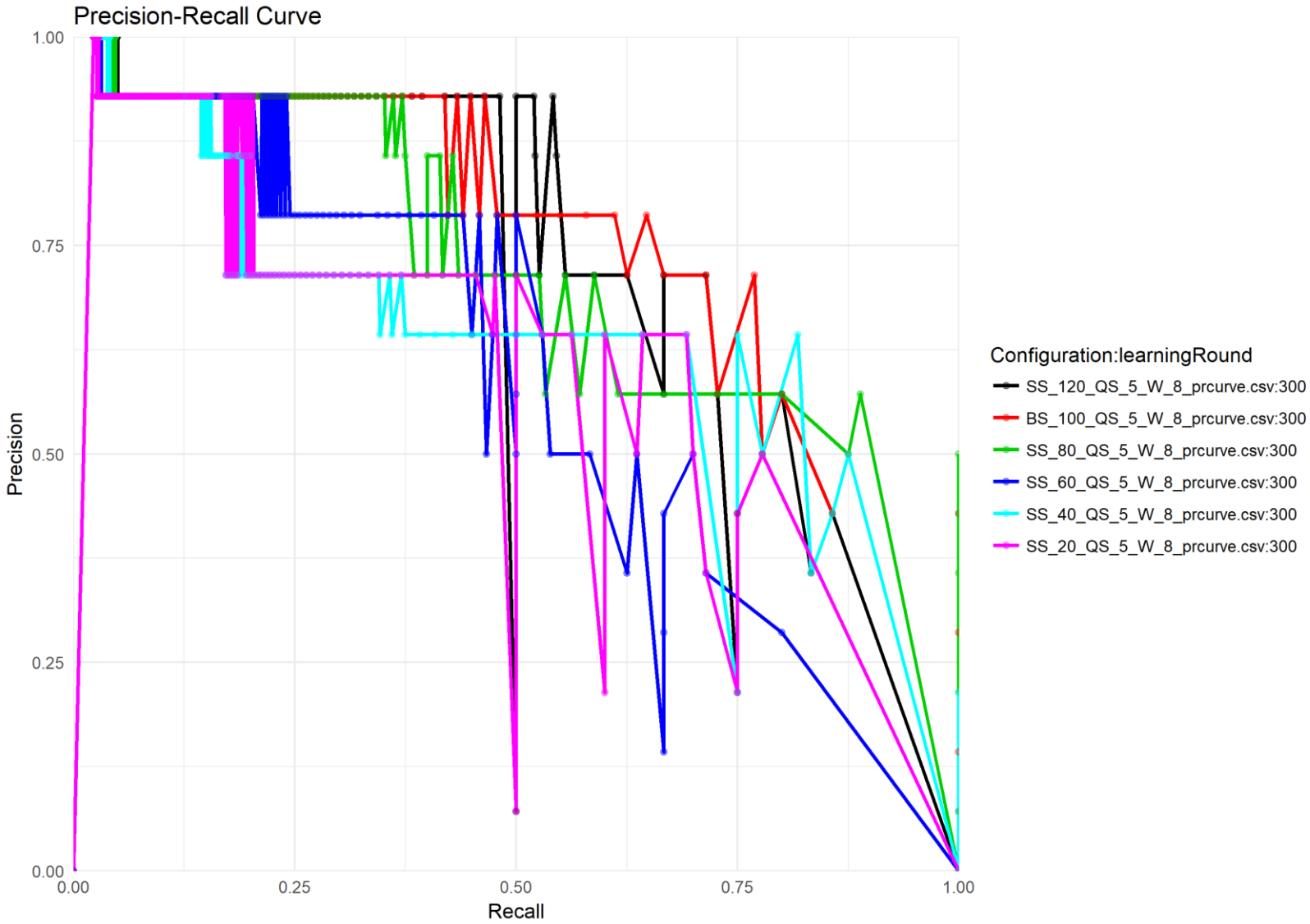
Comparison of Seed Set Size and Query Size Influence

- Weight true with factor 200 (w8)
  - Seed Set random sampled

Influence of Seed Set



Influence of Seed Set



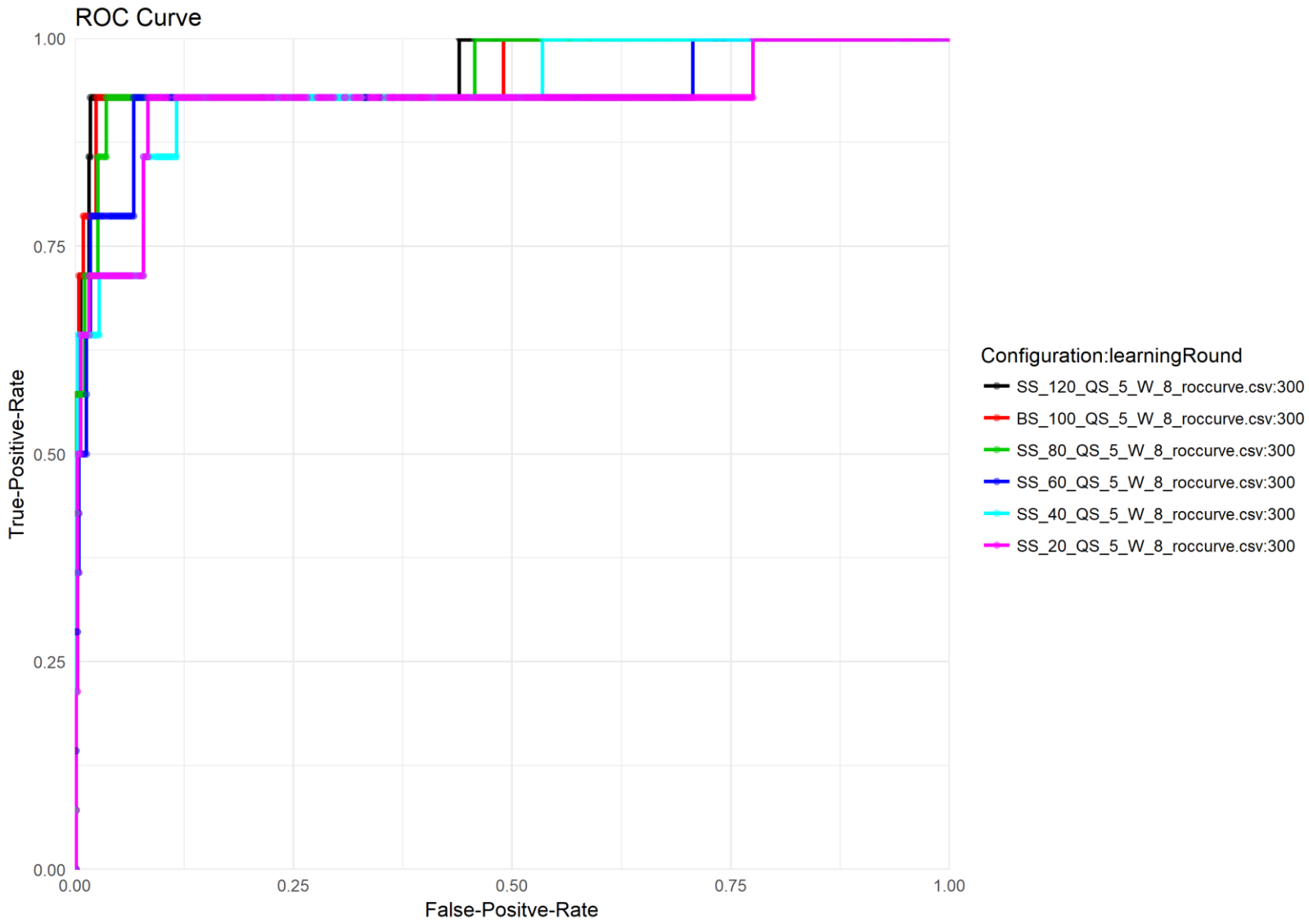
Reasons

- Small Random Sampled Seed Sets are often not representative
- Assumption BC 3% labeled as True:
  - P(no True in Seed Set with n=20) ≈ 54%

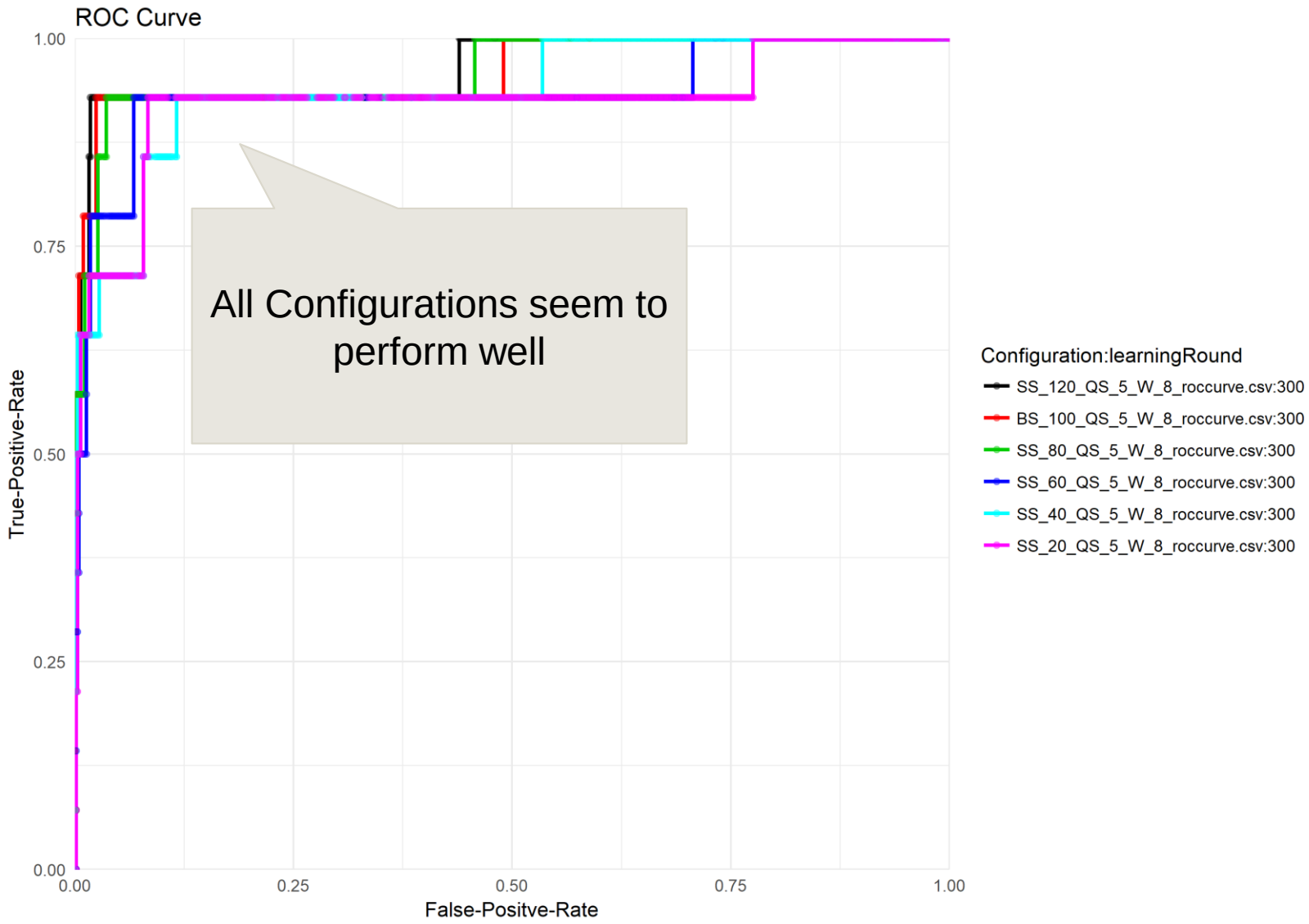
- Unbalanced classes can lead to missed class/cluster effect

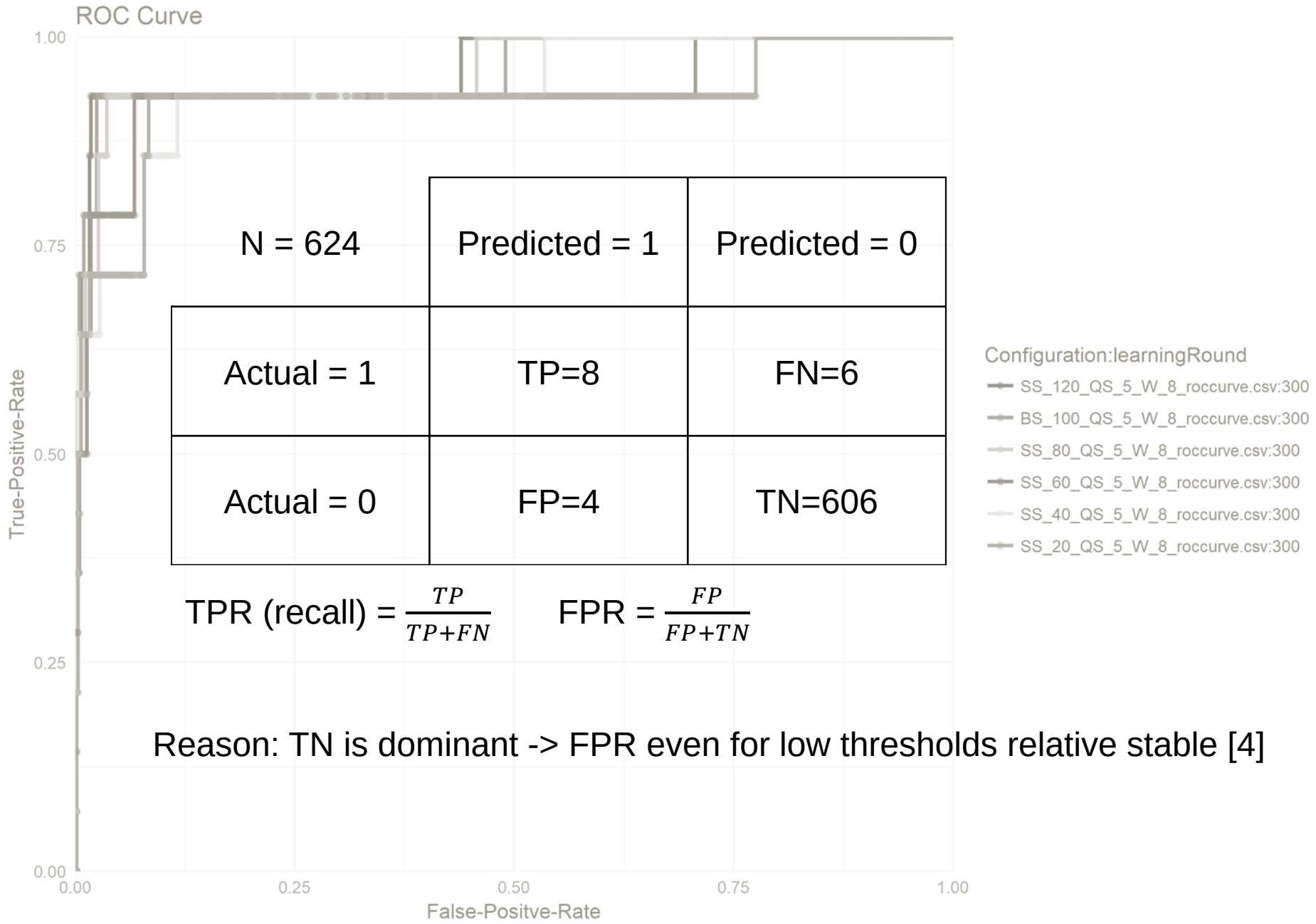
- => sufficient exploration phase during the seed set generation is crucial

ROC with unbalanced classes



ROC with unbalanced classes



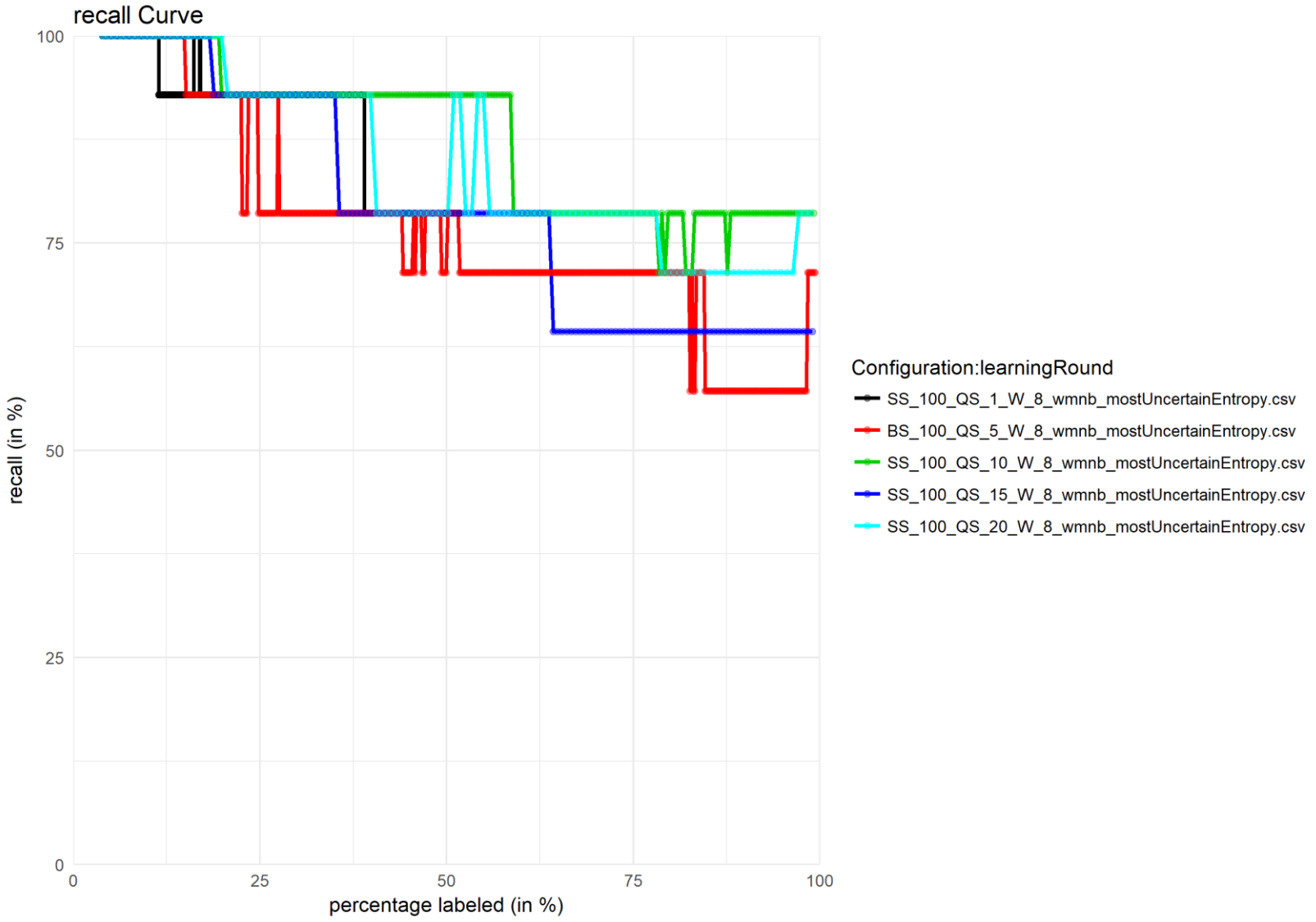


N = 624	Predicted = 1	Predicted = 0
Actual = 1	TP=8	FN=6
Actual = 0	FP=4	TN=606

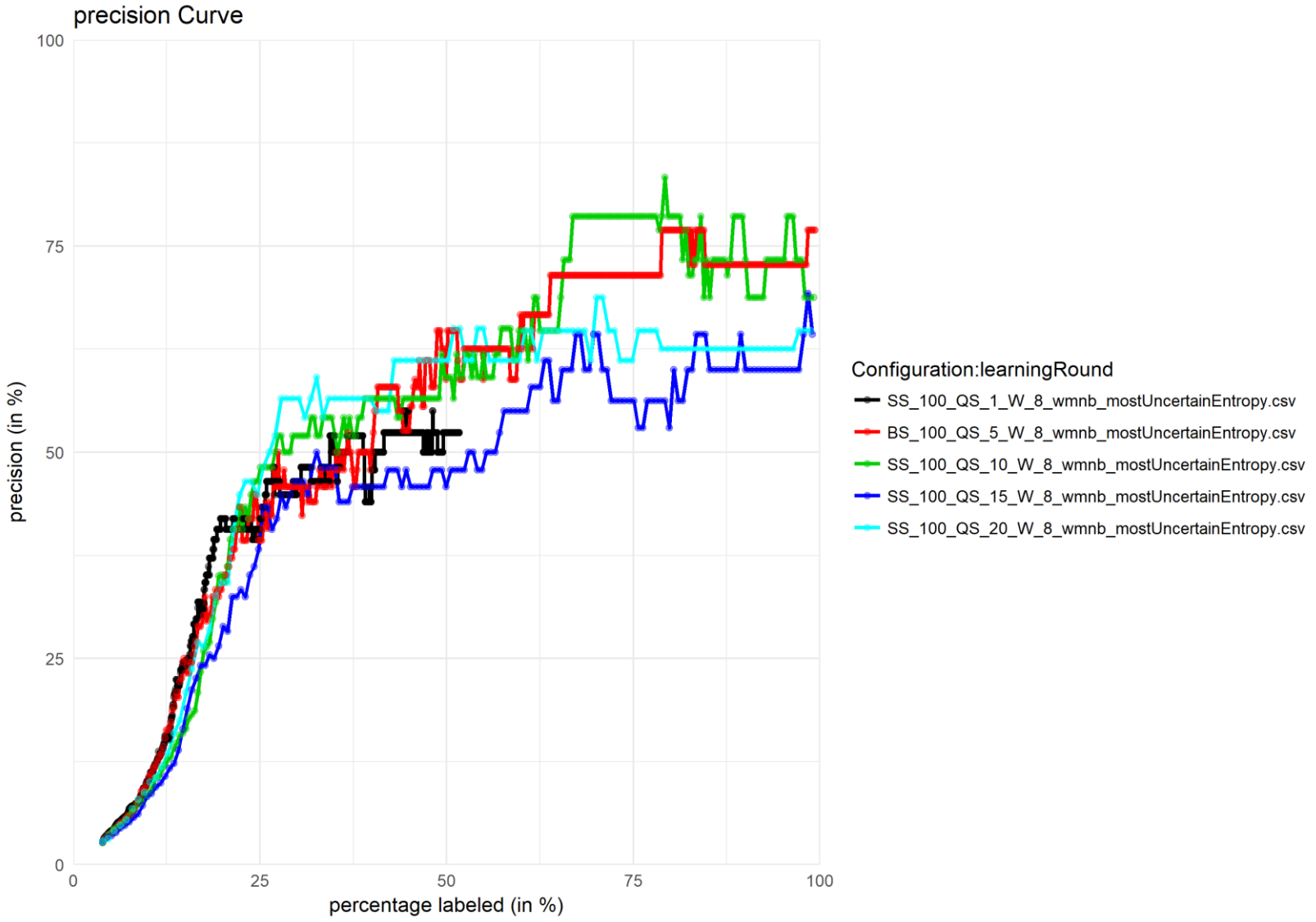
TPR (recall) = $\frac{TP}{TP+FN}$ FPR = $\frac{FP}{FP+TN}$

Reason: TN is dominant -> FPR even for low thresholds relative stable [4]

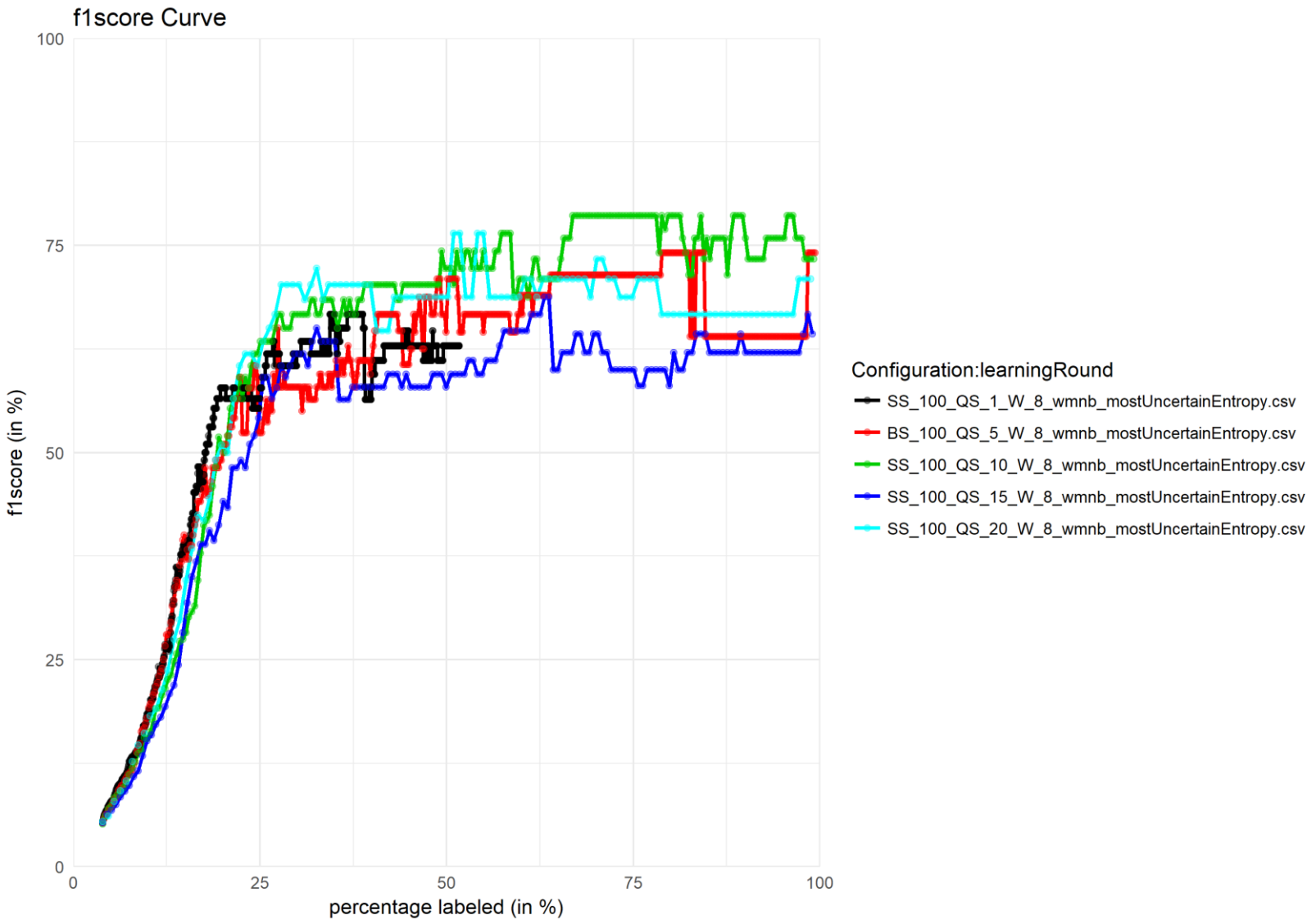
Influence of Query Size



Influence of Query Size

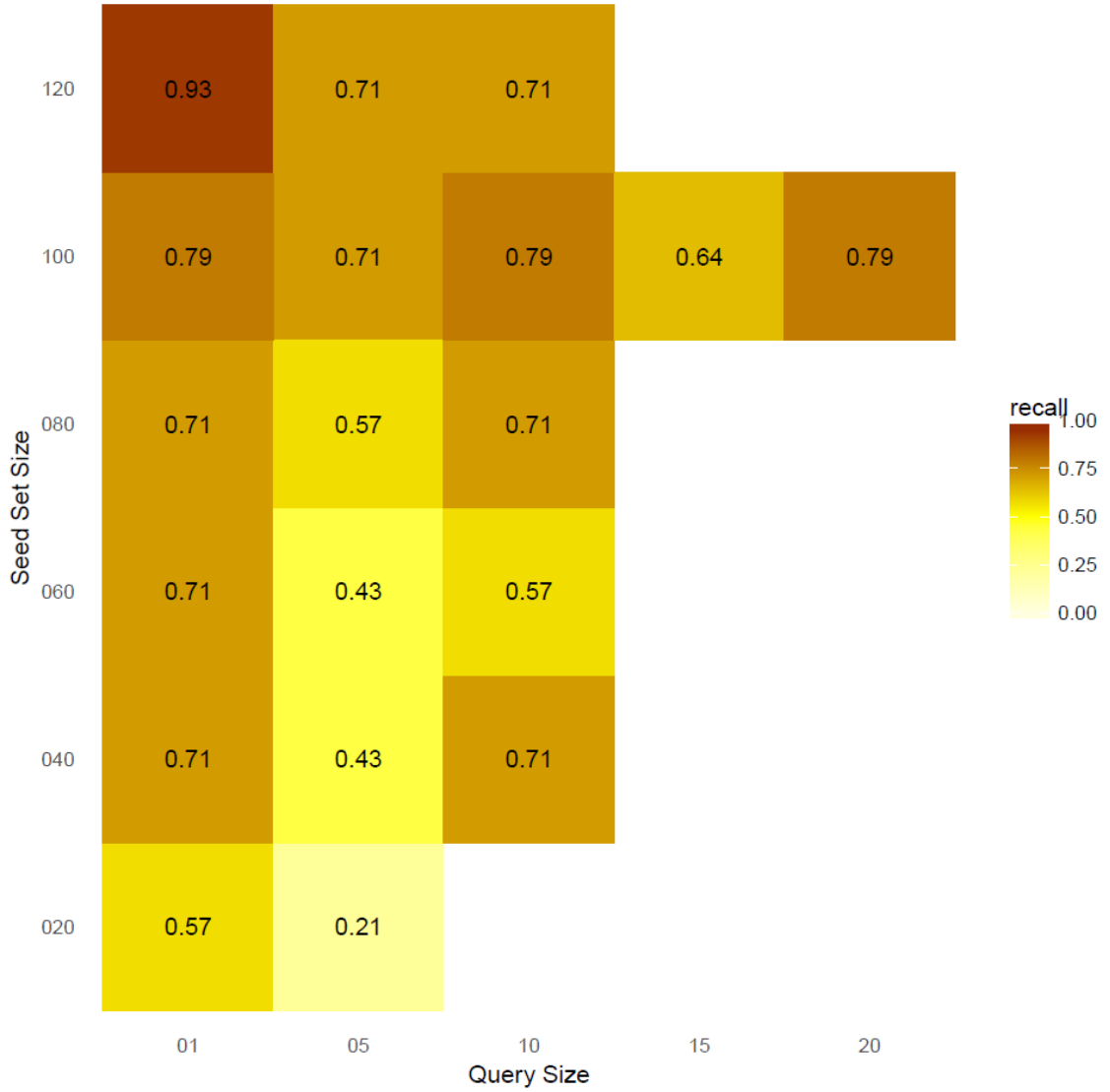


Influence of Query Size



Recall Heatmap

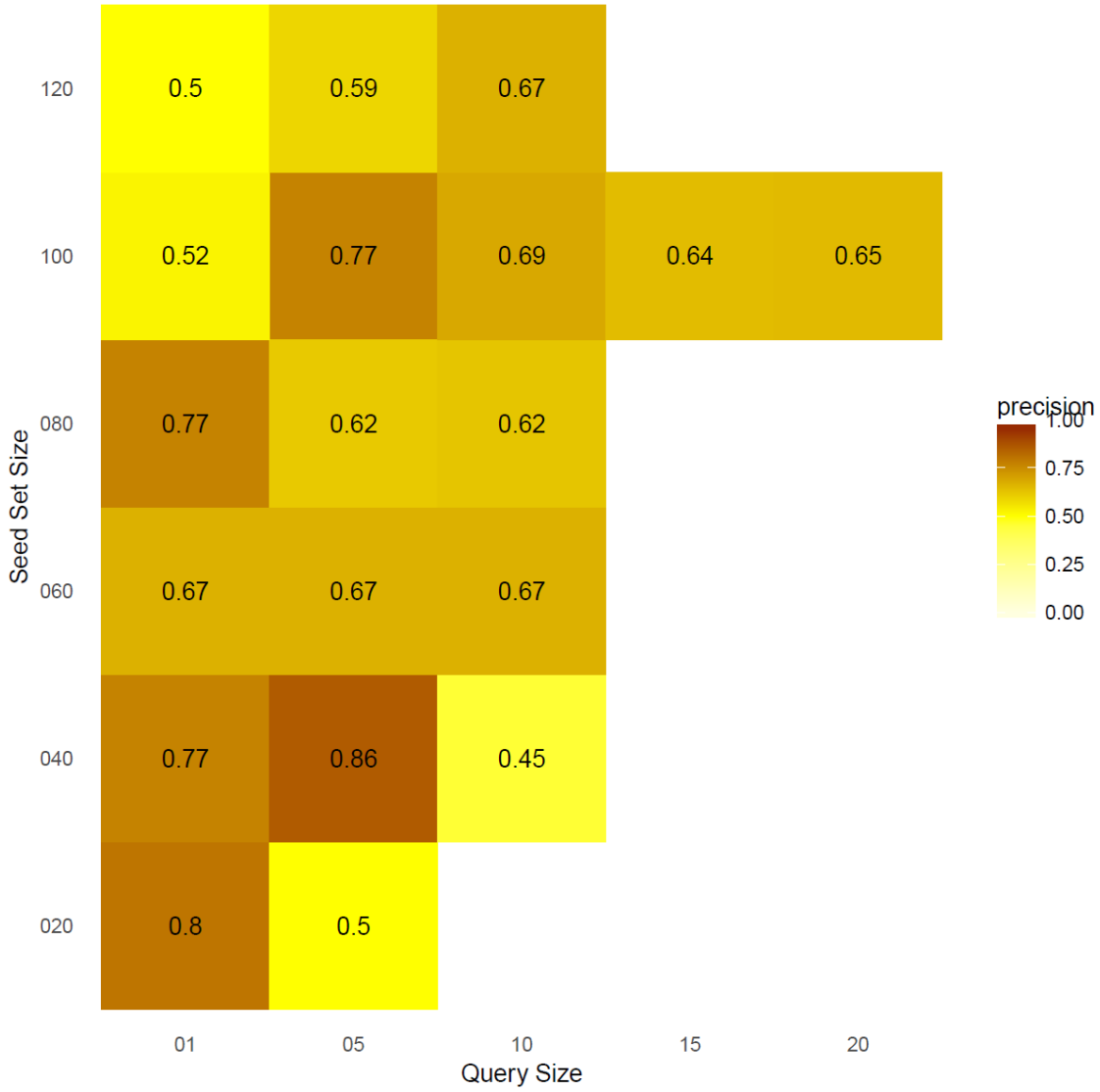
$$\text{Recall} = \frac{TP}{TP+FN}$$



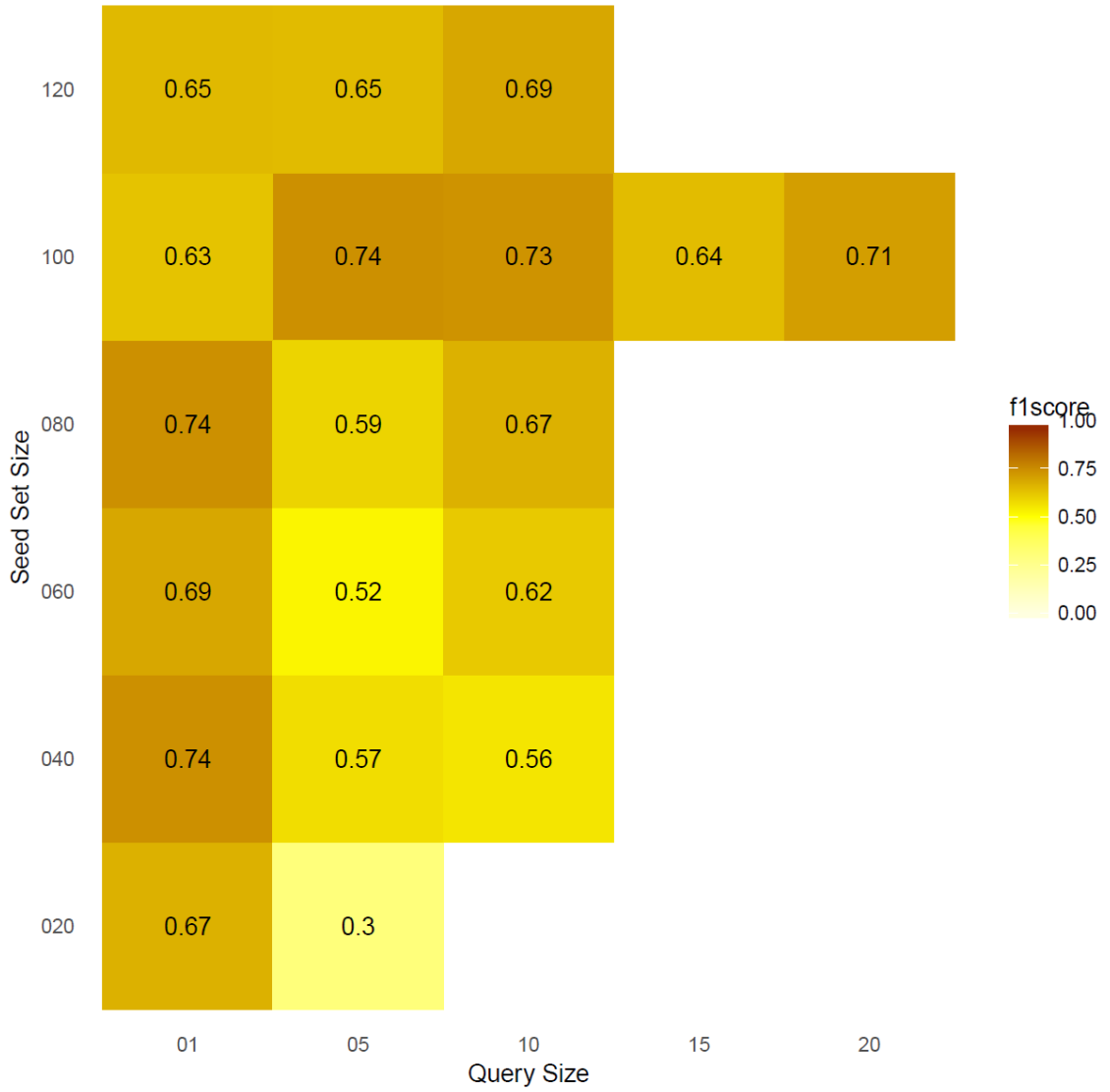
Precision Heatmap

$$\text{Precision} = \frac{TP}{TP+FP}$$

Low FP cause high precision even when



F1Score Heatmap



Demo

Conclusion

- Binary Classification is a promising approach for identifying predicates in judgments
- Seed Set Generation and Size is crucial

Limitations

- Evaluation with only one preprocessing configuration
- No Cross-Validation
- Strict Annotation rule misses some sentences

Future Work

- Evaluation of different preprocessing configurations in context of judgments
- Validate results with Cross-Validation
- Try different methods to encounter unbalenced classes
- Use a combination of AML and rule-based learning for a better class balance
- Using different predicates then Sentences like paragraphs or “n-grams” of sentences

Linus Boehm

Advisor: Ingo Glaser

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

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289. 17132
Fax +49.89.289.17136

matthes@in.tum.de
wwwmatthes.in.tum.de

References

(1) https://www.bundestag.de/blob/194870/7c8a01e16c98fc9c32ddb203d7bd88e0/gesetzgebung_wp18-data.pdf

(2) Gruner, Ronald H. "Anatomy of a lawsuit." (2008).

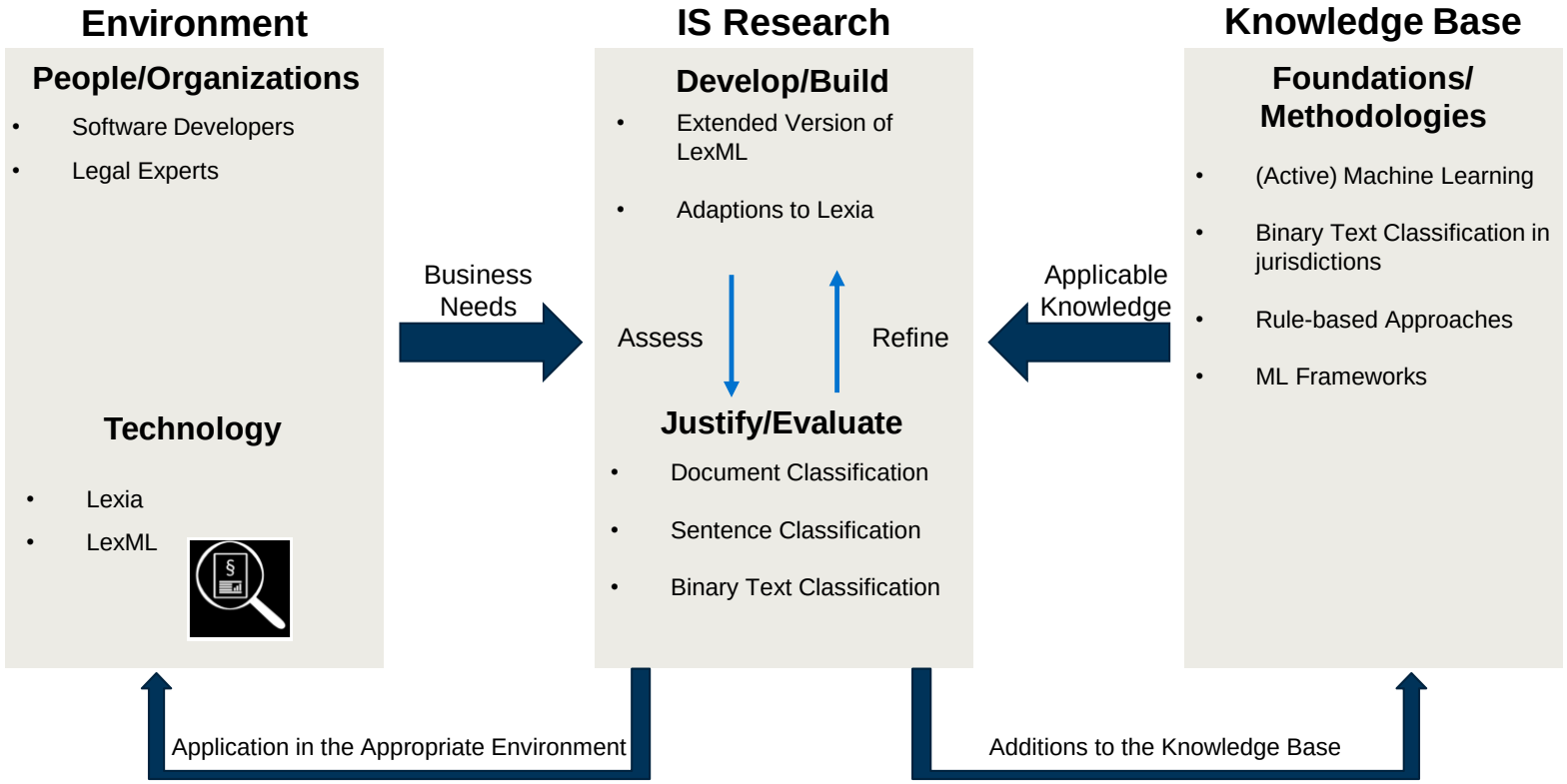
(3) Muhr, Johannes, Master Thesis, Design, Prototypical Implementation and Evaluation of an Active Machine Learning Service in the Context of Legal Text Classification (2017)

(4) https://www.kaggle.com/lct14558/imbalanced-data-why-you-should-not-use-roc-curve

Research Method

Design Science as most fitting method

- Focus is on the creation of an effective IT solution
- Description of the design problem
- Development an new IT artefact
- Strict evaluation of the IT artefact to measure the utility



Research Questions