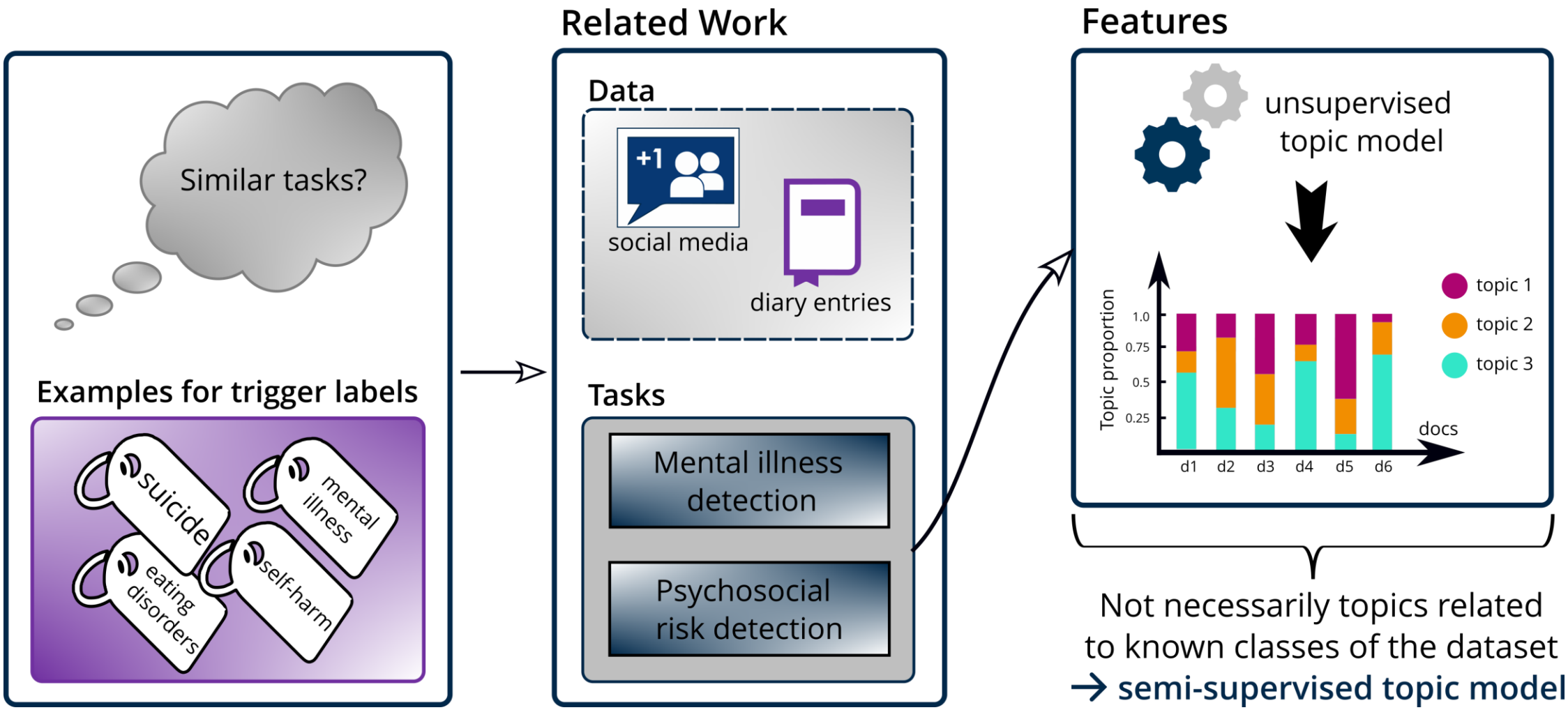
HOCHSCHULE MITTWEIDA
University of Applied Sciences

Trigger Detection with a Two Stage Topic Classifier

FoSIL at PAN'23

Jenny Felser

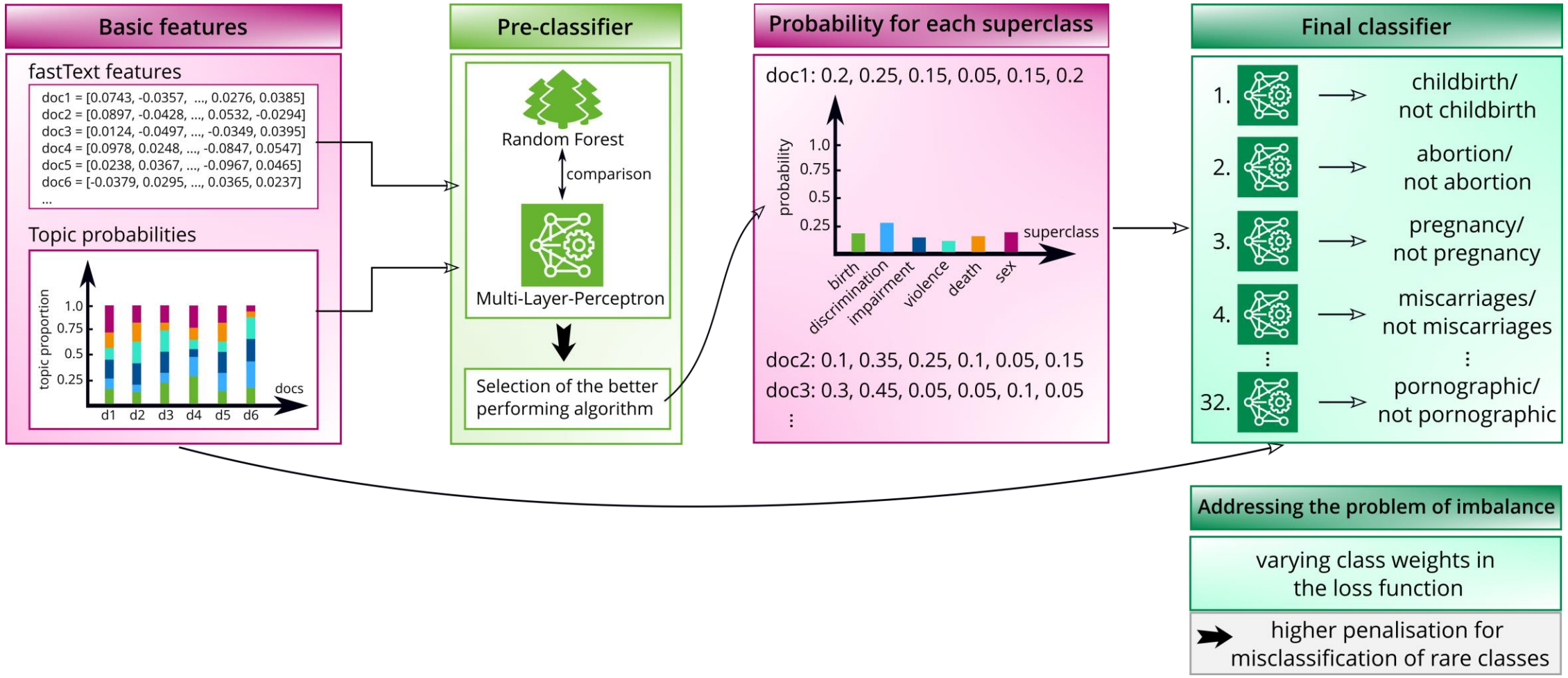
Related Work



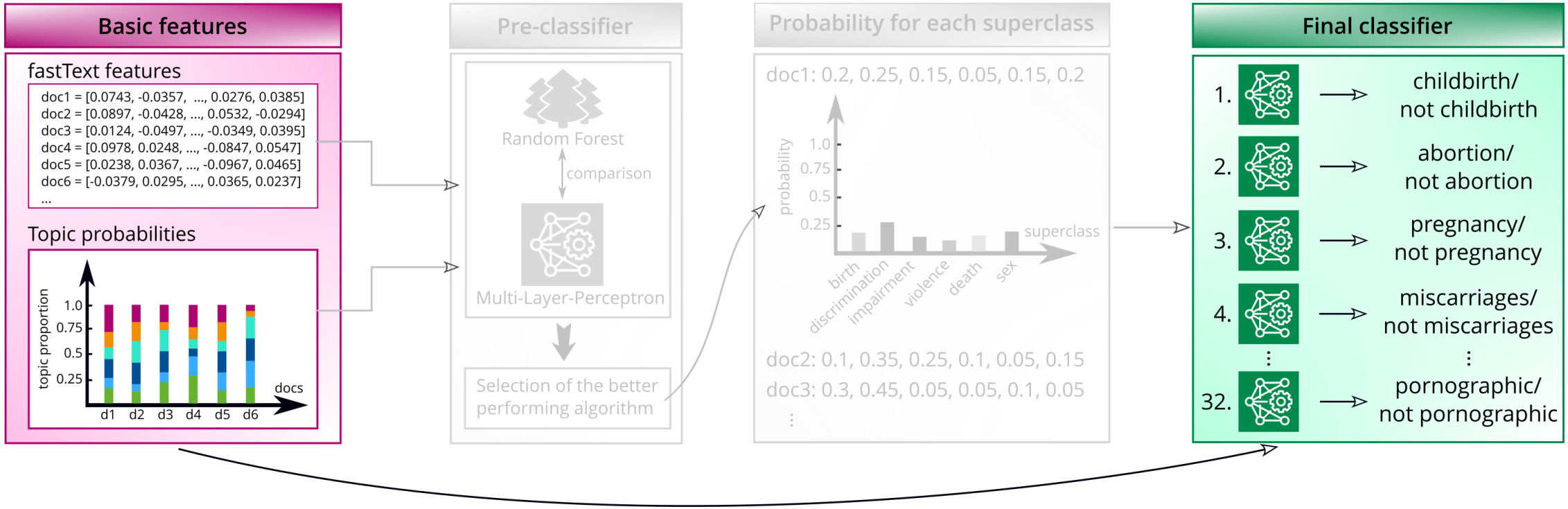
Manual creation of superclasses

Pre-classifier to predict the superclasses

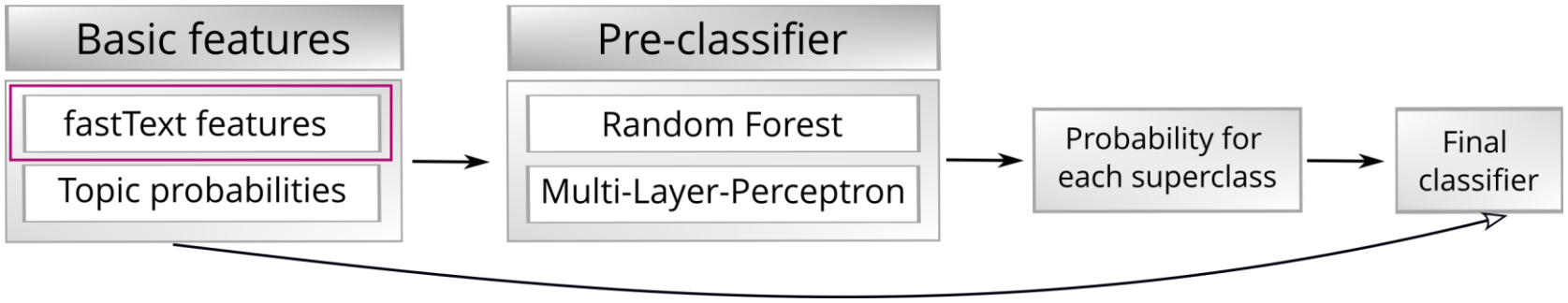
Two Stage Topic Classifier

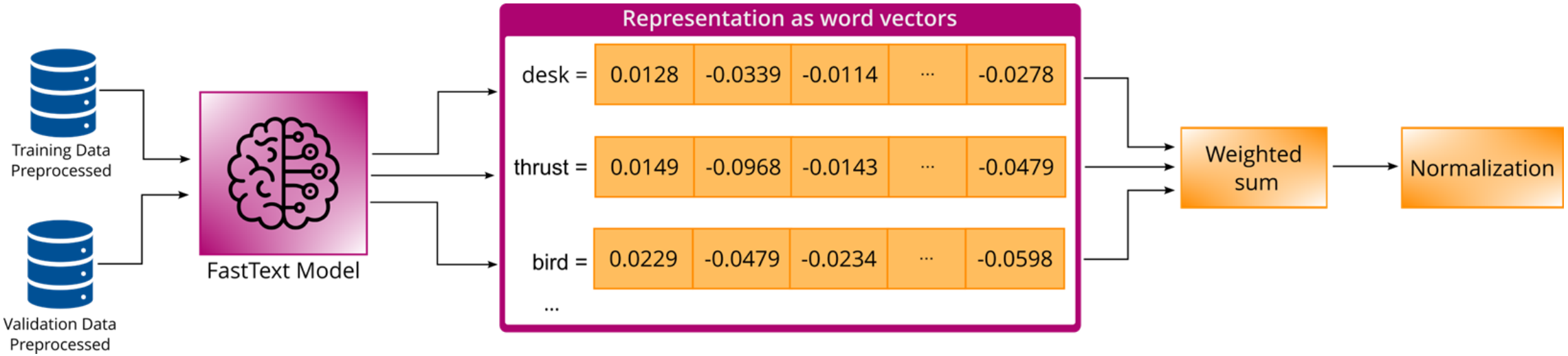


Baseline – One-stage classifier

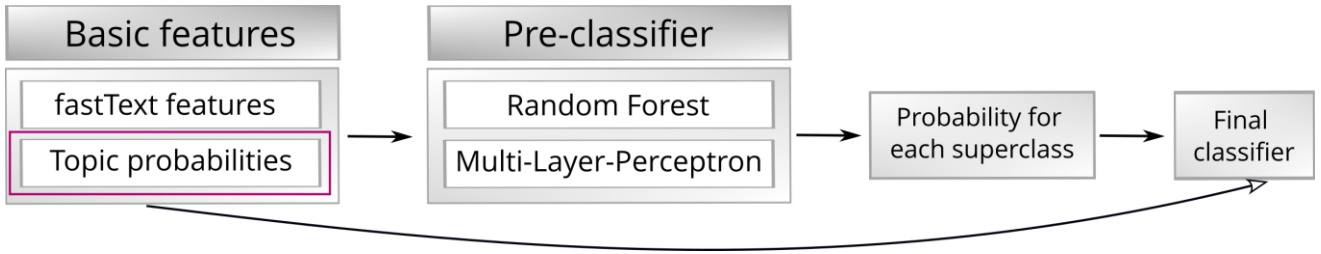


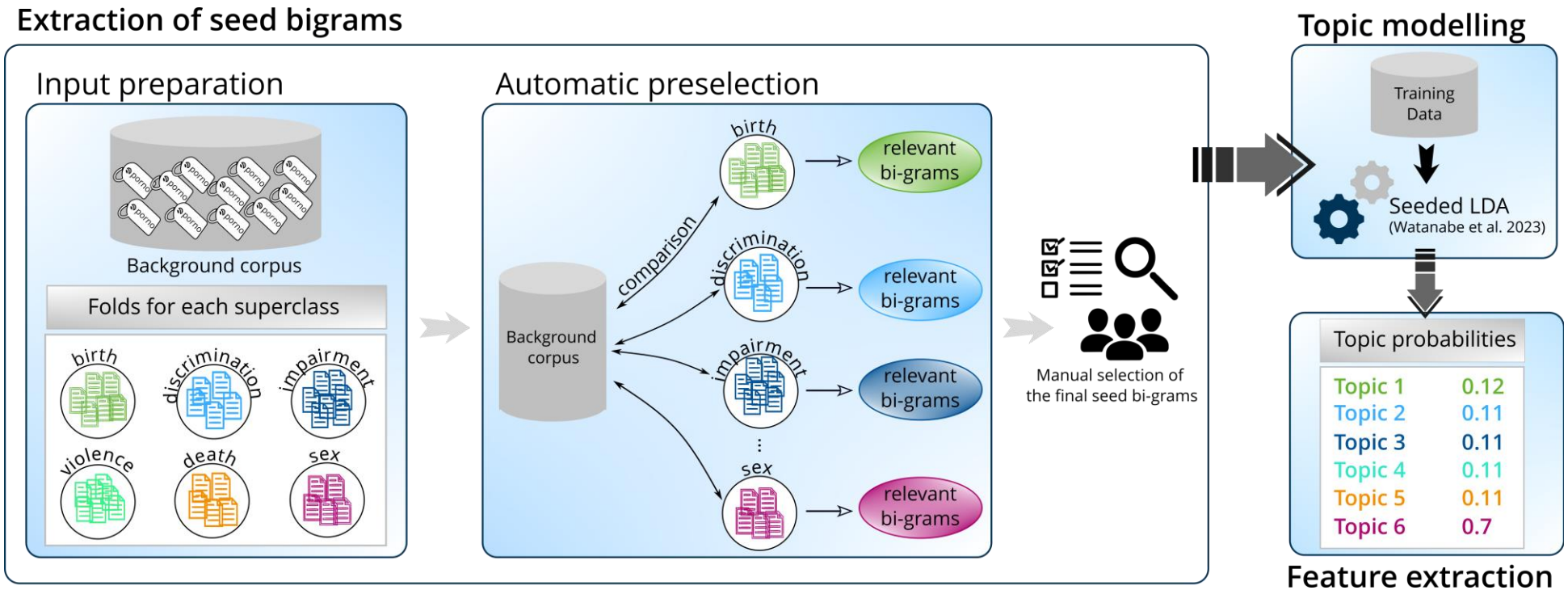
Creation of Document Vectors





Guided Topic Modeling

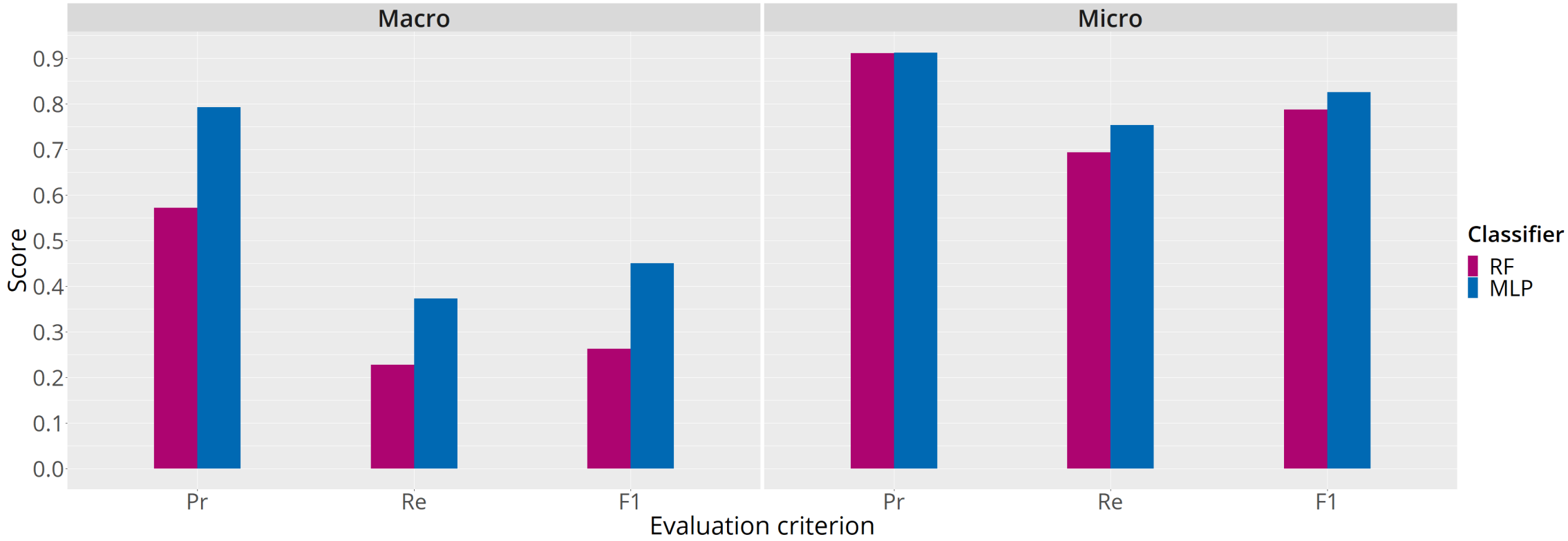




Guided Topic Modeling

Superclass	Seed bi-grams
birth/ pregnancy	baby shoe, pregnancy now, male carrier, may pregnant, baby gender, maternity pant
discrimination	officer bower, faggot word, mook knife, nash look, want fag, cecil belly, blood doll
impairment	spectrum disorder, know bipolar, binder feel, trauma holder, know autistic, purge night
violence	macaroni fuck, bubble bass, bitter root, psycho delic, robber girl
death	quirkless useless, funeral want, baby assassin, drooly loser, concept death, stealth shadow
sex	leg spread, now want, wrap around, around cock, finger inside, wrap lip, swirl tongue, mouth kiss, woundfucker number

Results – Pre-classifier



Selection of MLP for generating the superclass probabilities for the final classifier

Results - Final Classifier

Validation data

	Macro-Scores			Micro-Scores		
	Pr	**Re**	**F1**	**Pr**	**Re**	**F1**
MLP_S	0.0822	0.4530	0.1147	0.2751	0.7350	0.4003
MLP_E1	0.1135	0.6511	**0.1622**	0.2655	0.8214	0.4013
MLP_E2	0.1209	0.3925	0.1568	0.4451	0.6947	**0.5426**

Test data

	Macro-Scores			Micro-Scores		
	Pr	**Re**	**F1**	**Pr**	**Re**	**F1**
MLP_E1	0.113	0.631	**0.161**	0.266	0.818	0.402
MLP_E2	0.119	0.381	0.152	0.444	0.69	**0.54**

Findings

- Low macro F1-Scores
- Higher micro F1-Scores
- ➢ Some labels never predicted
- ➢ Highest recall at the cost of precision

- Better results through two-stage than one-stage approach

Future Work

Manually created superclasses

- Possible improvement through different grouping of trigger warnings

Better results by MLP_E2 than MLP_E1 due to the variation of weights

- Further parameter fine-tuning required

Probably dependencies between the trigger warnings

- Use of a classifier chain for the final classification

Thank you!

HOCHSCHULE MITTWEIDA
University of Applied Sciences

Jenny Felser
felser@hs-mittweida.de

Hochschule Mittweida | University of Applied Sciences
Technikumplatz 17 | 09648 Mittweida
Applied Computer Sciences & Biosciences | Cybercrime/ Cybersecurity

T +49 (0) 3727 58-1717

hs-mittweida.de