COVID-19 and Populism in Austrian News User Comments - A Machine Learning Approach

Master Defense

Supervisor: Julia Neidhardt
Co-Supervisor: Thomas Elmar Kolb

24.04.2023

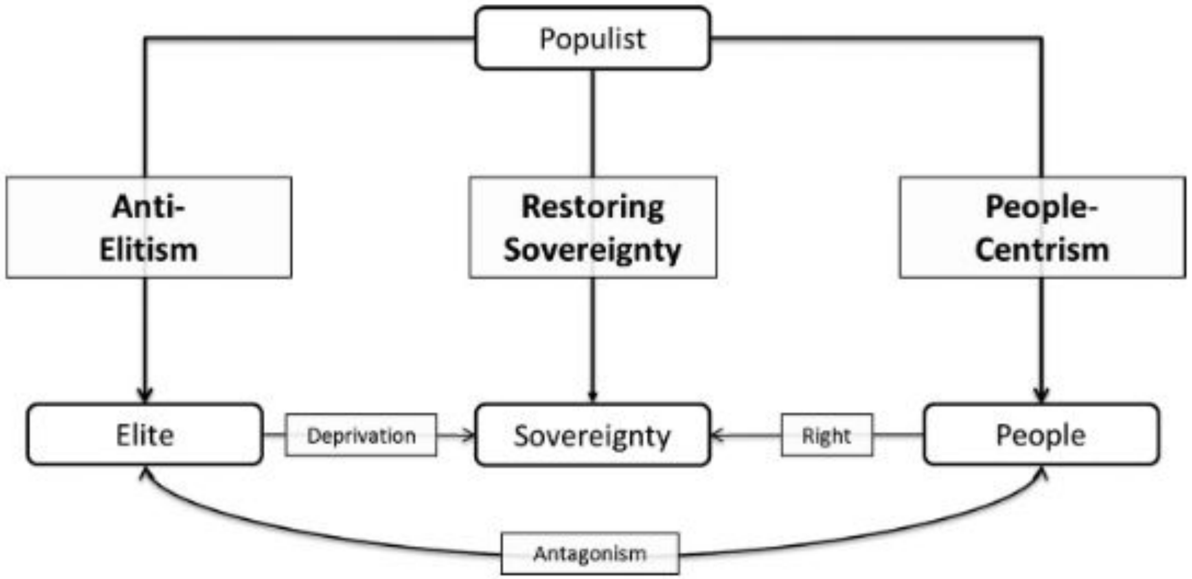
1. Motivation and Problem Statement

Ideational Definition

"An ideology that considers society to be ultimately separated into two homogeneous and antagonistic groups, 'the pure people' versus 'the corrupt elite', and which argues that politics should be an expression of the volonté générale (general will) of the people."

Cas Mudde 2004 – The Populist Zeitgeist

Operationalization of Populism



Motivation

- Can we find connections between COVID-19 and populist statements?
- Can we use machine learning to measure this on a large scale?
- Can we identify populism in short texts?

Problem Statement

- Research focuses on populism expressed by politicians
- Mostly qualitative or dictionary-based methods
- There is no annotated data for populist user comments
- Populism on Facebook has been investigated, but only from the beginning of the pandemic until May 2021
- Study uses a dictionary-based method tailored towards German social media content

TECHNISCHE UNIVERSITÄT WIEN

Research Questions

- **RQ1**: What is an appropriate ML model to improve the detection of populism in Austrian news comments?
- **RQ2a**: How does the COVID-19 crisis affect the amount of populist user comments?
- **RQ2b**: How does the topic of COVID-19 affect the amount of populist user comments posted during the crisis?
- **RQ3**: How does the amount of populist user comments under COVID-19-related articles evolve over time?

2. Data Analysis

Der Standard Data

- Kindly provided by Austrian daily newspaper *Der Standard*
- Filter articles for keywords (Corona, COVID, SARS)
- Three different samples: reference sample (January 2019 – December 2019), COVID-19 sample and non-COVID-19 sample (January 2020 - November 2021)

https://www.derstandard.at

Data: Article and user statistics

	Reference	COVID-19	Non-COVID-19
Total comments	9,250,241	14,016,469	16,152,559
Unique articles	49,421	14,135	68,642
Comments per article	187.2	991.6	235.3
Unique users	55,780	64,401	82,354
Comments per user	165.8	217.6	196.1

Data: Article and user statistics

	Reference	COVID-19	Non-COVID-19
Total comments	9,250,241	14,016,469	16,152,559
Unique articles	49,421	14,135	68,642
Comments per article	187.2	991.6	235.3
Unique users	55,780	64,401	82,354
Comments per user	165.8	217.6	196.1

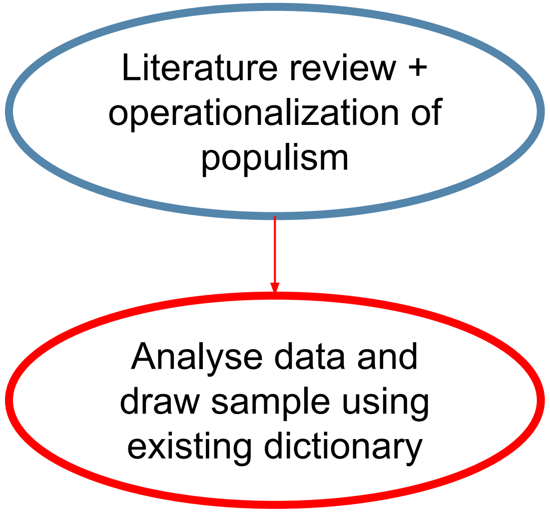
3. Methods

Workflow

Workflow

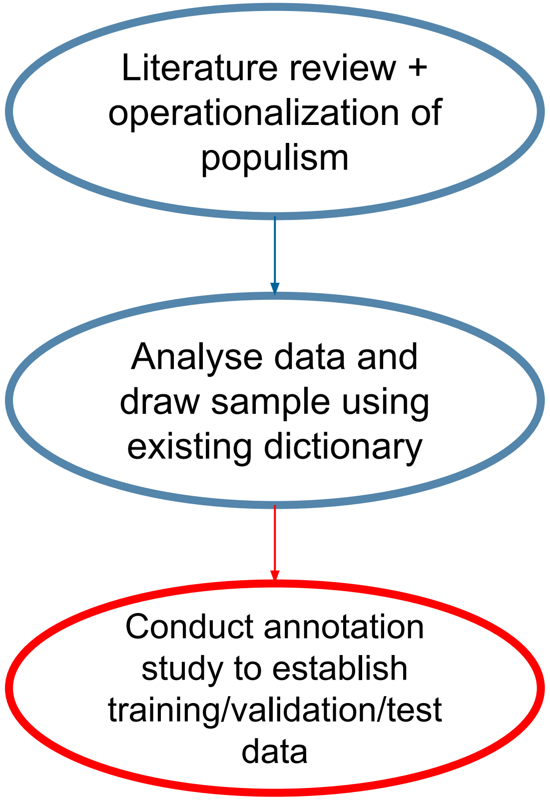
Literature review + operationalization of populism

Workflow

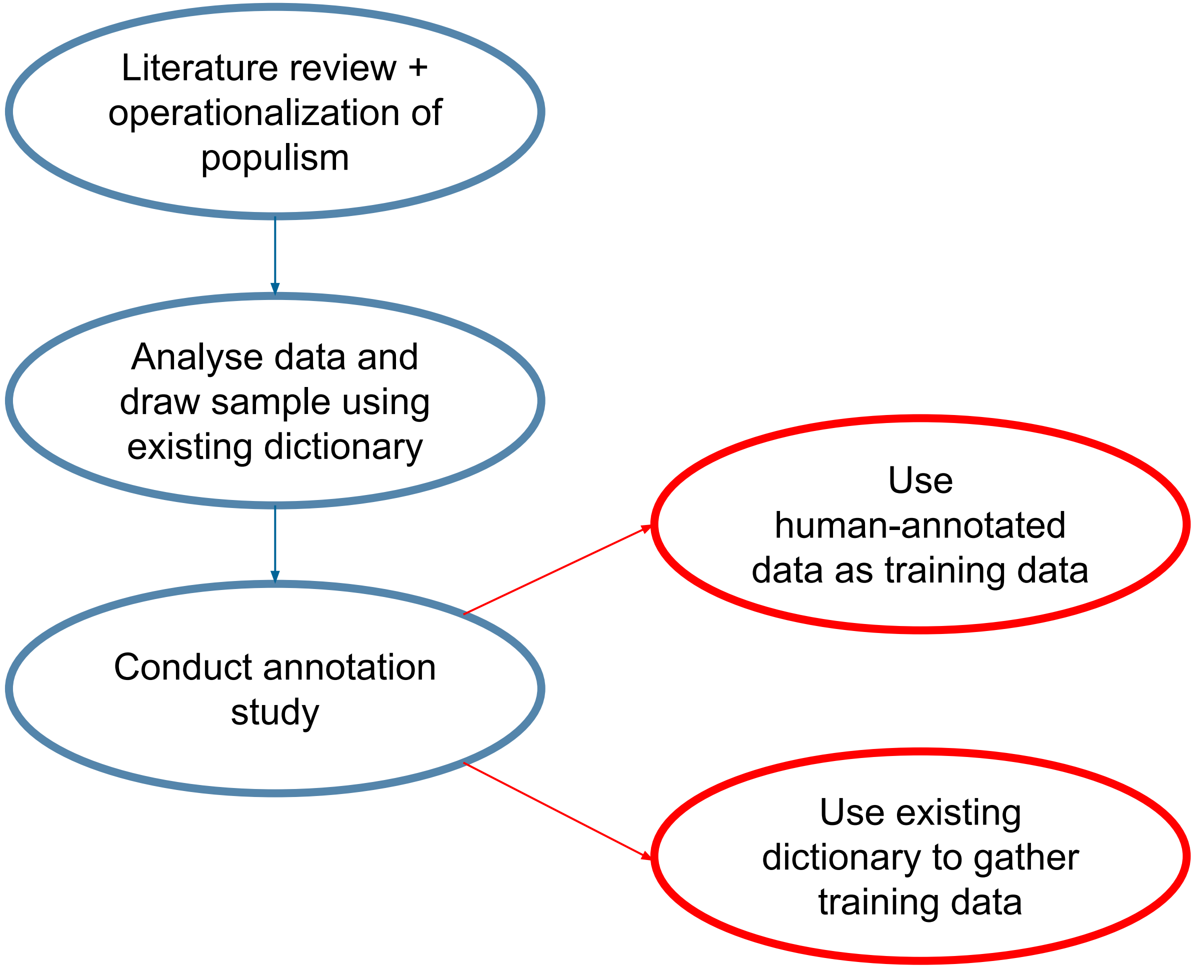


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary

Workflow


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary
Conduct annotation study to establish training/validation/test data

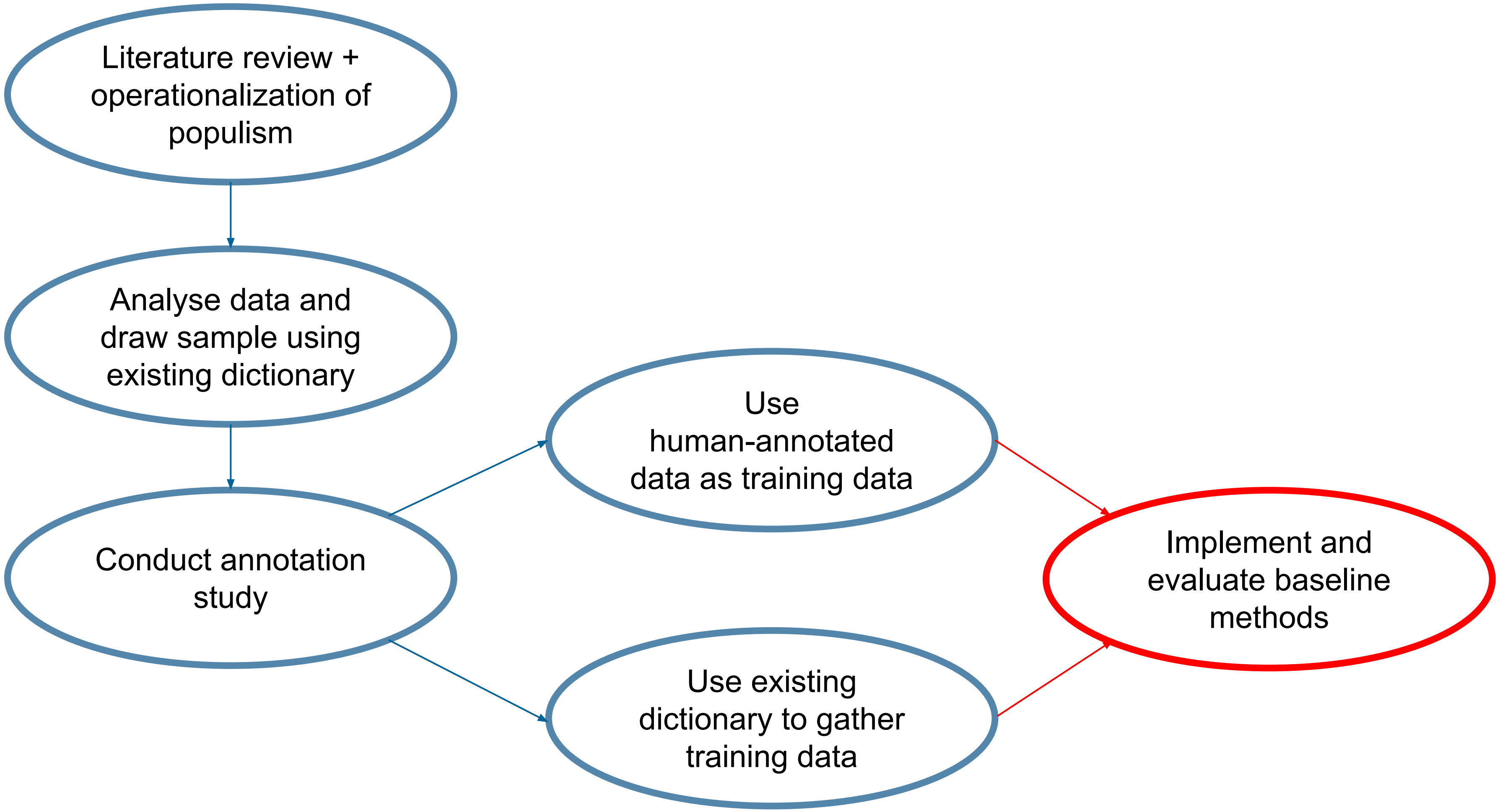
Workflow


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary
Conduct annotation study
Use human-annotated data as training data
Use existing dictionary to gather training data

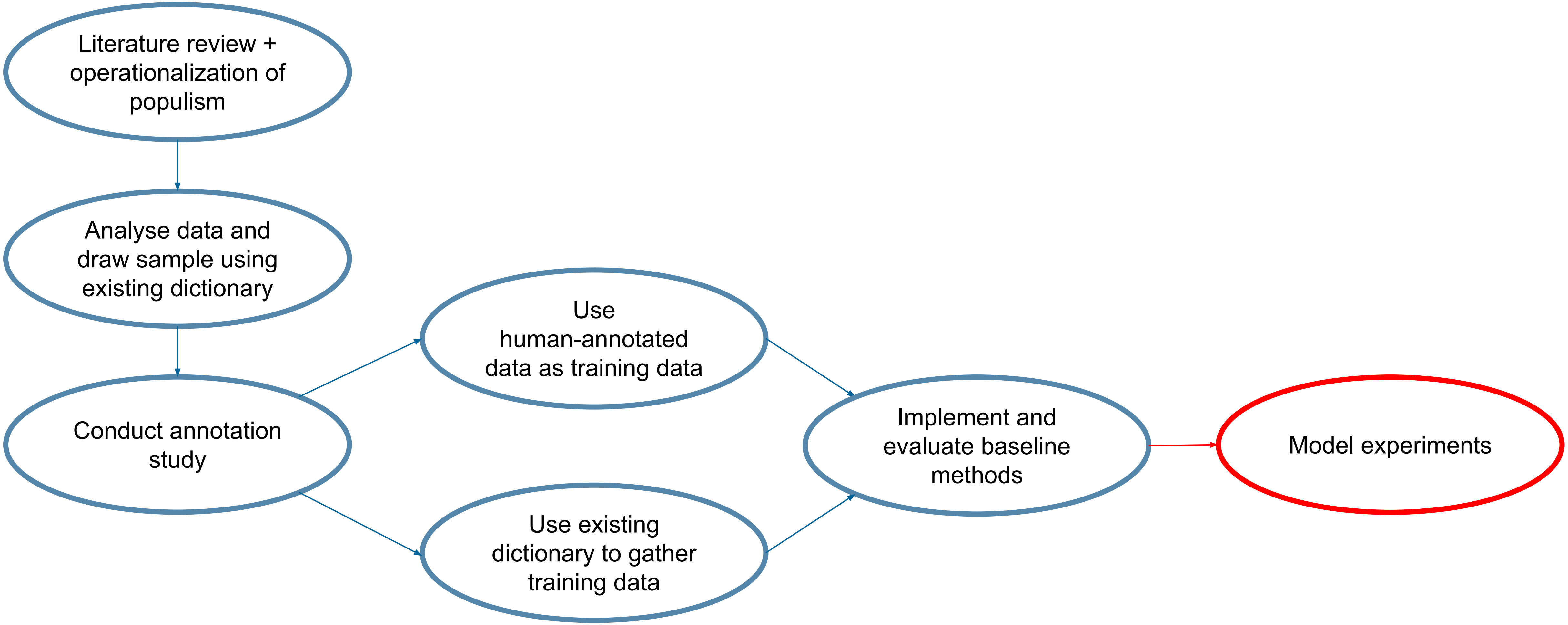
Approach

- Use the dictionary to automatically label a large amount of training data (12,000 comments) for fine-tuning of a BERT model
- Done before for sentiment analysis
- Use transformer model to overcome limitations of the dictionary
- Compare to performance of fine-tuning with a small human-annotated data (800 comments)

Workflow


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary
Conduct annotation study
Use human-annotated data as training data
Use existing dictionary to gather training data
Implement and evaluate baseline methods

Workflow


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary
Conduct annotation study
Use human-annotated data as training data
Use existing dictionary to gather training data
Implement and evaluate baseline methods
Model experiments

Approach

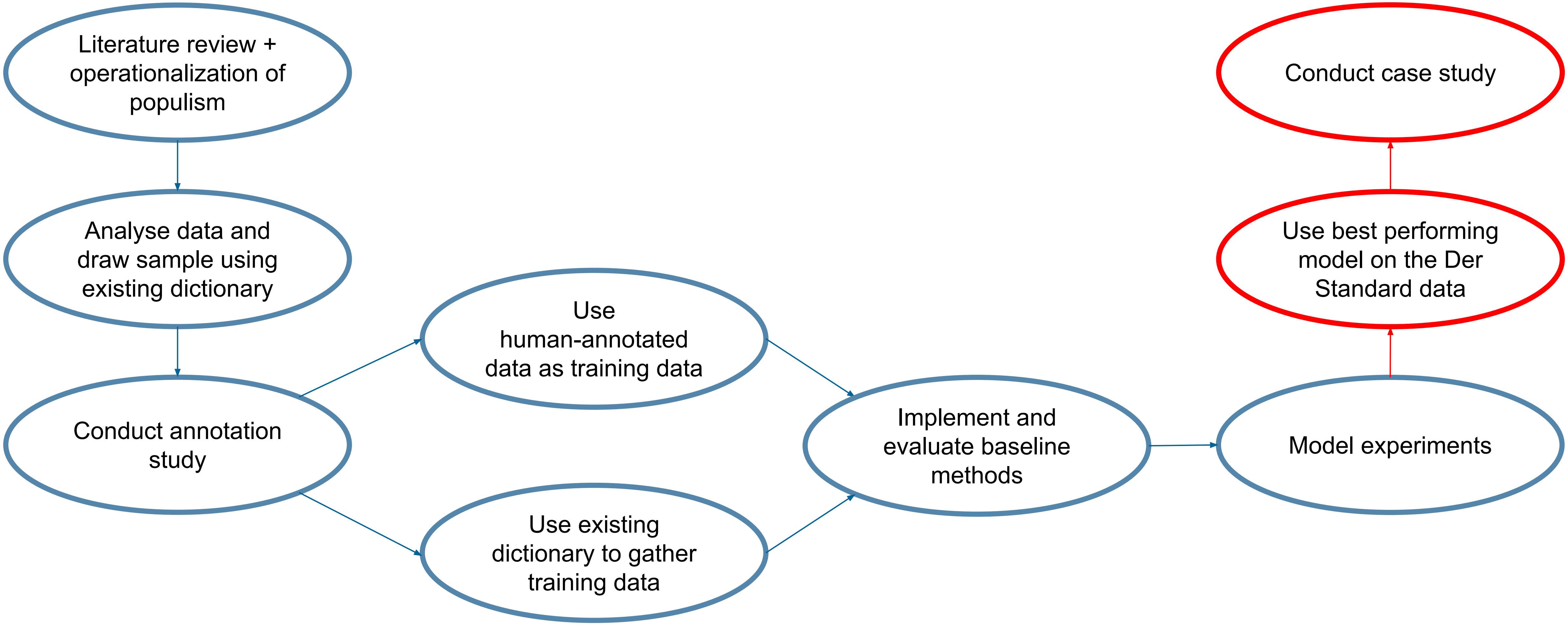
- Perform experiments with different pre-trained models and manipulations of the input data to reduce noise:
  - Initial cleaning (IC): remove HTML tags, URLs, non-ASCII characters, digits, single-letter words and multiple white spaces
  - Spelling correction (SC): use Hunspell and an Austrian German dictionary to correct spelling mistakes
  - Models: different pre-trained BERT models: bert-base-german-cased (M1) and deepset/gbert-base (M2)
- 4*4 different experiment setups

https://huggingface.co/bert-base-german-cased https://huggingface.co/deepset/gbert-base

Measurements

- Accuracy, Precision, Recall and F1
- Decisive criterion F1-score
- Each experiment setup is fine-tuned with five different seeds and average performance on the the same human annotated test set is reported

Workflow


Literature review + operationalization of populism
Analyse data and draw sample using existing dictionary
Conduct annotation study
Use human-annotated data as training data
Use existing dictionary to gather training data
Implement and evaluate baseline methods
Model experiments
Use best performing model on the Der Standard data
Conduct case study

Case Study

- Use best-performing model to predict all comments of the three samples
- Dependent variable: number of populist comments per article
- Explanatory variables:
  - **RQ2a**: Was the comment posted before or during the pandemic?
  - **RQ2b**: Was the comment posted under an article involving the topic of COVID-19?
  - **RQ3**: Days passed since the outbreak of the pandemic
- Dependent variable is a count variable and overdispersed -> negative binomial regression model
- Include logarithm of total number of comments of an article as an offset to the model

2019

vs.

COVID-19
+
Non-COVID-19
(2020/21)

Case Study

- Use best-performing model to predict all comments of the three samples
- Dependent variable: number of populist comments per article
- Explanatory variables:
  - **RQ2a**: Was the comment posted before or during the pandemic?
  - **RQ2b**: Was the comment posted under an article involving the topic of COVID-19?
  - **RQ3**: Days passed since the outbreak of the pandemic
- Dependent variable is a count variable and overdispersed -> negative binomial regression model
- Include logarithm of total number of comments of an article as an offset to the model

COVID-19

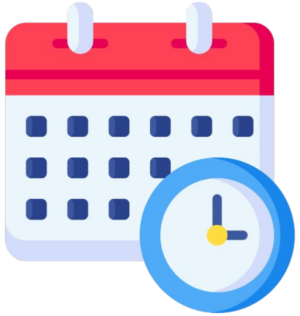
vs.

Non-COVID-19

Case Study

- Use best-performing model to predict all comments of the three samples
- Dependent variable: number of populist comments per article
- Explanatory variables:
  - **RQ2a**: Was the comment posted before or during the pandemic?
  - **RQ2b**: Was the comment posted under an article involving the topic of COVID-19?
  - **RQ3**: Days passed since the outbreak of the pandemic
- Dependent variable is a count variable and overdispersed -> negative binomial regression model
- Include logarithm of total number of comments of an article as an offset to the model

COVID-19



Case Study

- Use best-performing model to predict all comments of the three samples
- Dependent variable: number of populist comments per article
- Explanatory variables:
  - **RQ2a**: Was the comment posted before or during the pandemic?
  - **RQ2b**: Was the comment posted under an article involving the topic of COVID-19?
  - **RQ3**: Days passed since the outbreak of the pandemic
- Dependent variable is a count variable and overdispersed -> negative binomial regression model
- Include logarithm of total number of comments of an article as an offset to the model

4. Results

Annotation Study

Agreement on populist labels

	Agreement (Krippendorff's α)
Anti-Elitism	0.79
People-Centrism	0.54
People-Sovereignty	0.72
Populism	0.79

- Three annotators per comment
- Multiple choice for populist motives
- Presence of one motive indicates populism
- Final label based on majority vote

Final labels after majority vote

	Final Sample
Anti-Elitism	271
People-Centrism	47
People-Sovereignty	22
Populism	297
None	903

Annotation Study

Agreement on populist labels

	Agreement (Krippendorff's α)
Anti-Elitism	0.79
People-Centrism	0.54
People-Sovereignty	0.72
Populism	0.79

- Three annotators per comment
- Multiple choice for populist motives
- Presence of one motive indicates populism
- Final label based on majority vote

Final labels after majority vote

	Final Sample
Anti-Elitism	271
People-Centrism	47
People-Sovereignty	22
Populism	297
None	903

Annotation Study

Agreement on populist labels

	Agreement (Krippendorff's α)
Anti-Elitism	0.79
People-Centrism	0.54
People-Sovereignty	0.72
Populism	0.79

- Three annotators per comment
- Multiple choice for populist motives
- Presence of one motive indicates populism
- Final label based on majority vote

Final labels after majority vote

	Final Sample
Anti-Elitism	271
People-Centrism	47
People-Sovereignty	22
Populism	297
None	903

Annotation Study

Agreement on populist labels

	Agreement (Krippendorff's α)
Anti-Elitism	0.79
People-Centrism	0.54
People-Sovereignty	0.72
Populism	0.79

- Three annotators per comment
- Multiple choice for populist motives
- Presence of one motive indicates populism
- Final label based on majority vote

Final labels after majority vote

	Final Sample
Anti-Elitism	271
People-Centrism	47
People-Sovereignty	22
Populism	297
None	903

IC: initial cleaning
SC: spelling correction
M1: bert-base-german-cased
M2: deepset/gbert-base

Experiments with 12,000 Dictionary-Annotated Training Samples

	IC + M1 +SC	IC + M1	IC + M2	IC + M2 + SC	M1 + SC	M2 + SC	M1	M2
Accuracy	**0.63**	**0.63**	**0.63**	**0.63**	**0.63**	**0.63**	**0.63**	**0.63**
Precision	0.38	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	0.38
Recall	0.87	0.86	**0.88**	0.86	0.87	0.87	0.87	0.87
F1	0.53	0.53	**0.54**	0.54	**0.54**	**0.54**	**0.54**	0.53

- Similar performance for all experiment setup
- No improvement over the dictionary

IC: initial cleaning
SC: spelling correction
M1: bert-base-german-cased
M2: deepset/gbert-base

Experiments with Human Annotated Training Data

	IC + M1 +SC	IC + M1	IC + M2	IC + M2 + SC	M1 + SC	M2 + SC	M1	M2
Accuracy	0.81	0.8	0.79	0.76	0.81	**0.82**	0.79	**0.82**
Precision	0.62	0.57	0.56	0.55	0.61	**0.63**	0.56	0.61
Recall	0.64	**0.79**	**0.79**	0.68	0.74	0.61	0.73	0.64
F1	0.61	**0.66**	0.65	0.56	**0.66**	0.62	0.62	0.62

- Best setups IC + M1 and SC + M2
- Decision for higher recall and less computational cost
- Improved precision and F1 compared to the dictionary

TECHNISCHE UNIVERSITÄT WIEN

CD Lab RecSys

Baseline: Populism Dictionary
Best large: best experiment setup with 12,000 dictionary-annotated training samples
Best small: best experiment setup with 800 human-annotated training samples
Single best: single best-performing model used for the case study

Comparison: Baseline vs experiments

	Baseline	Best large	Best small	Single best
Accuracy	0.7	0.63 (-0.07)	0.8 (+0.1)	**0.84 (+0.14)**
Precision	0.45	0.39 (-0.06)	0.57 (+0.12)	**0.63 (+0.17)**
Recall	**0.88**	**0.88 (+/-0)**	0.79 (-0.09)	0.86 (-0.02)
F1	0.59	0.54 (-0.05)	0.66 (0.07)	**0.72 (+0.13)**

RQ1: What is an appropriate ML model to improve the detection of populism in Austrian news comments?

⇨ Fine-tuning the bert-base-german cased model with initial cleaning and human annotated training data

Test statistics of negative binomial regression model

	Coefficient	Standard error	z	p-value	[0.025	0.975]
RQ2a	-0.0778	0.007	-11.783	<0.001	-0.091	-0.065

Case Study

- **RQ2a: How does the COVID-19 crisis affect the number of populist user comments?**
  - The average article published during the pandemic attracts ~7.8% less populist comments than an article published during 2019

Test statistics of negative binomial regression model

	Coefficient	Standard error	z	p-value	[0.025	0.975]
RQ2a	-0.0778	0.007	-11.783	<0.001	-0.091	-0.065
RQ2b	0.17	0.01	16.449	<0.001	0.15	0.19

Case Study

- **RQ2a: How does the COVID-19 crisis affect the number of populist user comments?**
  - The average article published during the pandemic attracts ~7.8% less populist comments than an article published during 2019
- **RQ2b: How does the topic of COVID-19 affect the number of populist user comments posted during the crisis?**
  - The average article dealing with the topic of COVID-19 attracts ~17% more populist comments than an article with another topic

Case Study

Test statistics of negative binomial regression model

	Coefficient	Standard error	z	p-value	[0.025	0.975]
RQ2a	-0.0778	0.007	-11.783	<0.001	-0.091	-0.065
RQ2b	0.17	0.01	16.449	<0.001	0.15	0.19
RQ3	-0.0006	<0.0001	-10.16	<0.001	-0.001	0

- **RQ2a: How does the COVID-19 crisis affect the number of populist user comments?**
  - The average article published during the pandemic attracts ~7.8% less populist comments than an article published during 2019
- **RQ2b: How does the topic of COVID-19 affect the number of populist user comments posted during the crisis?**
  - The average article dealing with the topic of COVID-19 attracts ~17% more populist comments than an article with another topic
- **RQ3: How does the amount of populist comments under COVID-19-related articles evolve over time?**
  - No noticeable effect of the duration of the crisis on the number of populist comments

5. Conclusion

Contribution

- First annotated sample for populist user comments in the German language
- Proposed model outperforms the state-of-the-art of populism detection in Austrian news comments
- Measured effects of the crisis on the number of populist user comments compared to 2019
- Performed analysis of populist user comments on a larger amount of data and a greater timespan that partially includes the fourth COVID-19 wave

Future Work

- Incorporate information about mentioned elites as features
- Analyse the populist content of specific users and user networks
- Include more news sources
- Focus on „pandemic populism“
- Label more data

Discussion