GHENT UNIVERSITY
FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Reducing labeling effort with active learning in sentiment classification

Presented by: Gatien Laurier
Thesis supervisor: Prof. dr. Matthias Bogaert

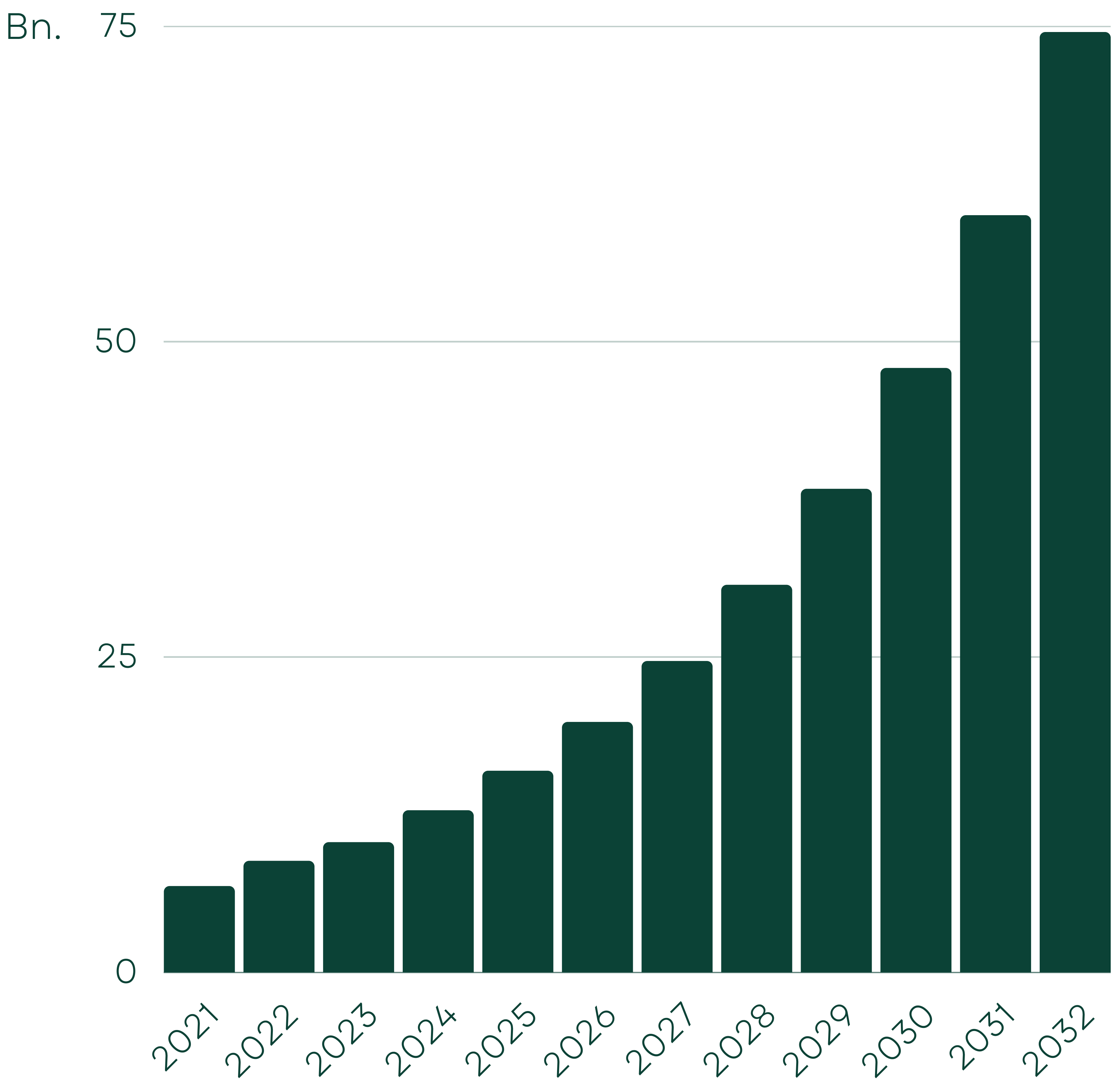
Data labeling market

24% CAGR

may experience some challenges in meeting the growing demand for labeled data

How to do better with less?

Source: Fact.MR



GHENT UNIVERSITY

Objective Statement

01

Reducing labeling effort

02

Accounting for non-expert labelers

GHENT UNIVERSITY

Prediction model

RobBERT

A Dutch RoBERTa-based Language Model

- Just like the well-known GPT, robBert is a large language model
- Is pre-trained on general Dutch language
- Can be fine-tuned for any classification task
- Less labeled data needed to train an accurate model
- If trained from scratch, loads of unlabeled data needed and some labeled data

Shares this characteristic with...

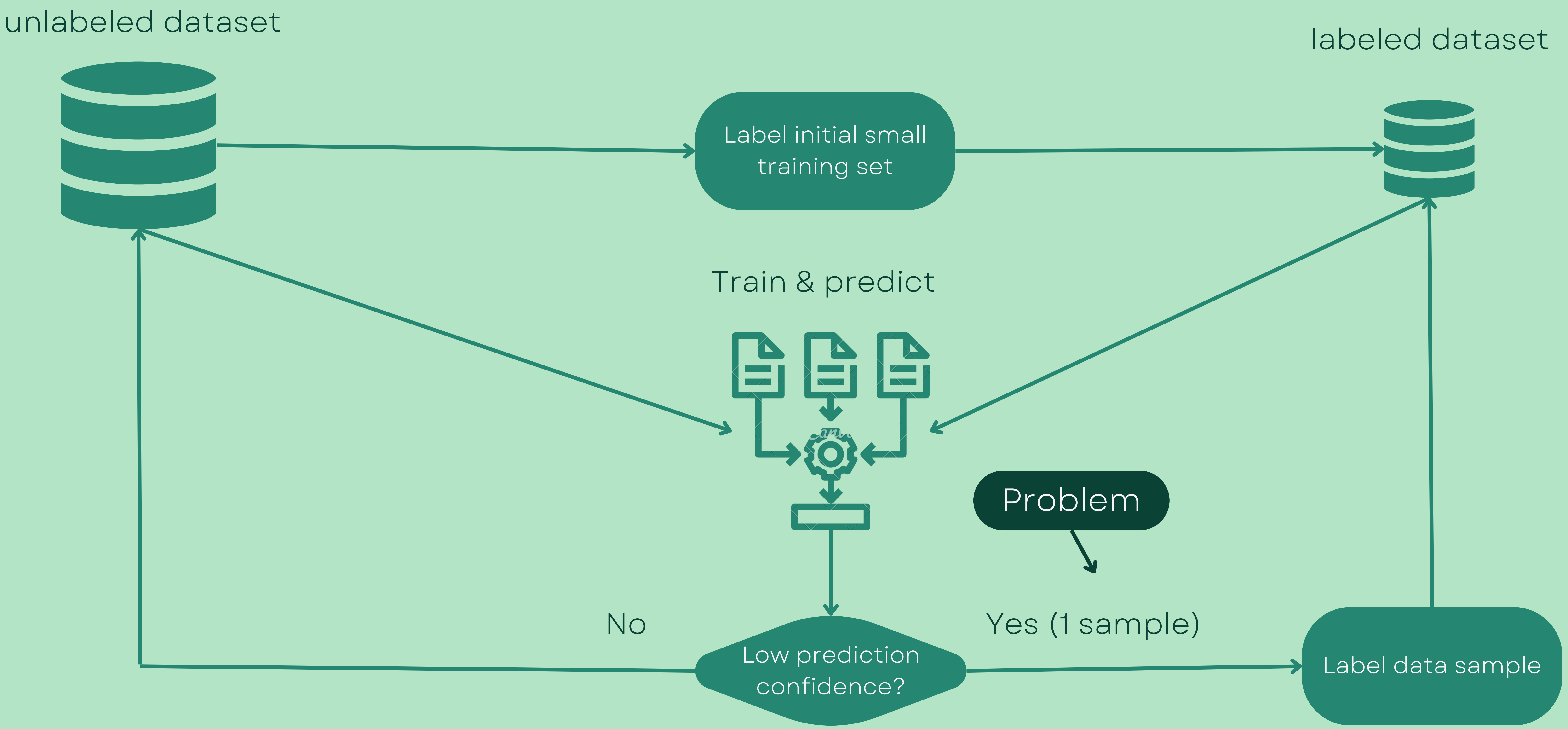
GHENT UNIVERSITY

Active learning

- Is an alternative approach to traditional way of training a prediction model
- Performs very well in situations where cost of collecting data is relatively low to labeling data
- This technique has been widely demonstrated to be highly effective under certain conditions in the literature.
- Can greatly reduce the required number of data points.

Objective 1: Active learning

GHENT UNIVERSITY



Problem - solution proposal

Entropy score (uncertainty)

Select points that current model is uncertain about

×

Density score

Avoid outliers, select samples in dense areas

×

Dissimilarity score

Avoid sampling similar points as they contain same information

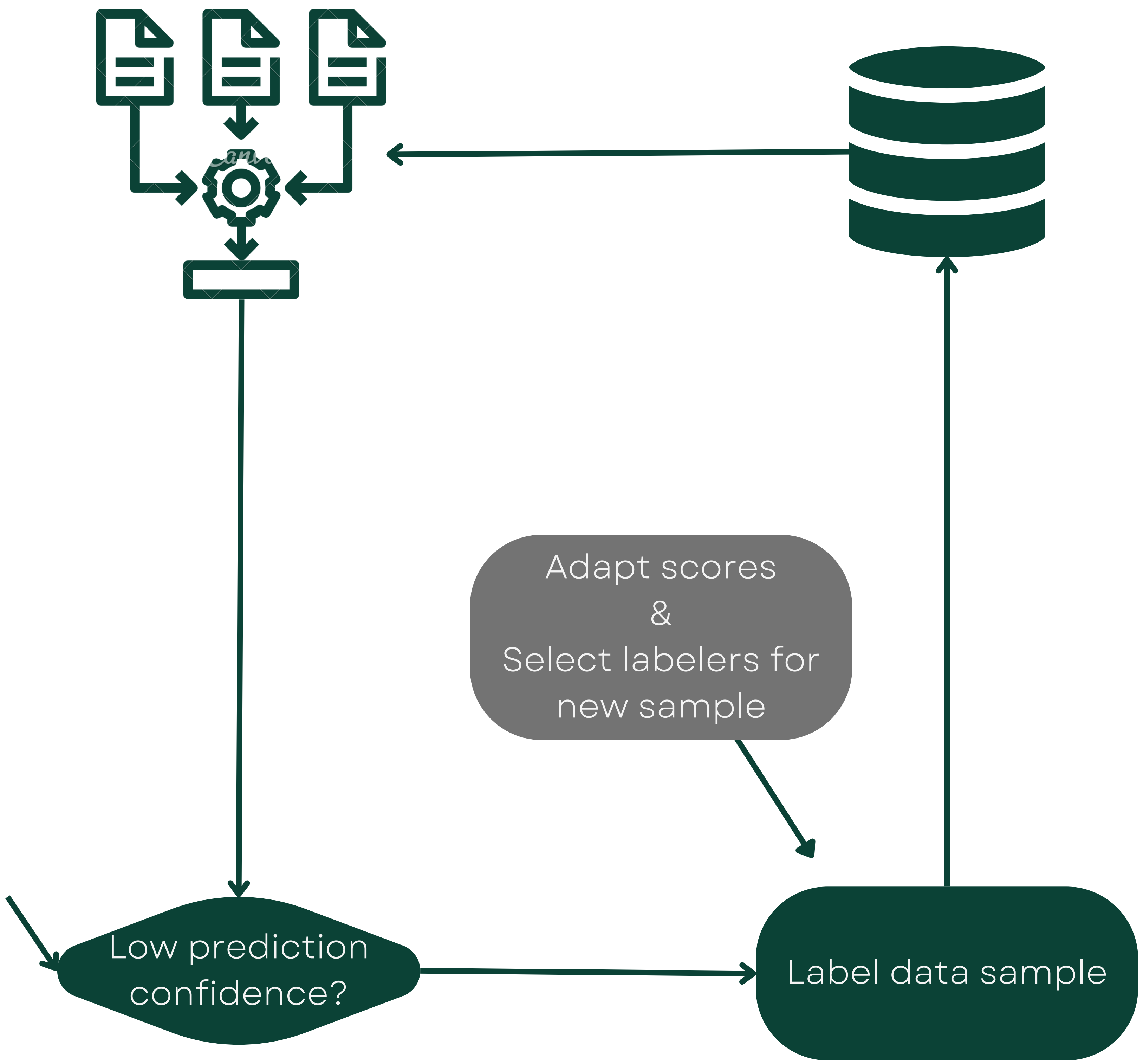
=

Final score

Select samples with highest score

Objective 2: proposed solution

- Making use of active learning structure to give score to annotators - majority voting
- Once score below threshold, labeler gets excluded
- However strong assumption made: annotation quality independent of input
- We will try to tackle by using topic modeling on unlabeled training pool - BERTopic



Source: Donmez, P., Carbonell, J. G., & Schneider, J. (2009). Efficiently learning the accuracy of labeling sources for selective sampling. Knowledge Discovery and Data Mining. https://doi.org/10.1145/1557019.1557053

GHENT UNIVERSITY

Example

	Thema 1: Covid-19	Thema 2: verkeer
	3/3	2/2
	3/3	2/2
	3/3	2/2

GHENT UNIVERSITY

Threshold: 75%

Example

	Thema 1: Covid-19	Thema 2: verkeer
	Negative 4/4	~~Neutral 2/3~~
	Negative 4/4	Positive 3/3
	Neutral 3/4	Positive 3/3

Objective 2 - non-expert labelers

Less low-quality annotators

After some iterations, labelers will fall below threshold and excluded

Quality scores dependent on input

Some annotators will be better in annotating some topics

Exploration phase

To avoid sanctioning too heavy too early

Less labeling

Only use labelers for a datapoint which are accurate

Adversarial learning

- Proven that using sampling on both sides of decision boundary is effective.
- Unsupervised learning is not precise
- Performing attacks on neural network by replacing similar words
- HotFlip: White-Box Adversarial Examples for Text Classification
- 2 for 1

Source: Ebrahimi, J., Rao, A., Lowd, D., & Dou, D. (2018). HotFlip: White-Box Adversarial Examples for Text Classification. https://doi.org/10.18653/v1/p18-2006

GHENT UNIVERSITY

Example

Class: Neutral

@BelgianGreenClub @MDiependaele Inderdaad, en het is zo dat iedereen zijn eigen verantwoordelijkheid heeft.

Class: Negative

@BelgianGreenClub @MDiependaele Inderdaad, en het is alleen dat iedereen zijn eigen verantwoordelijkheid heeft.

Example

Class: Neutral

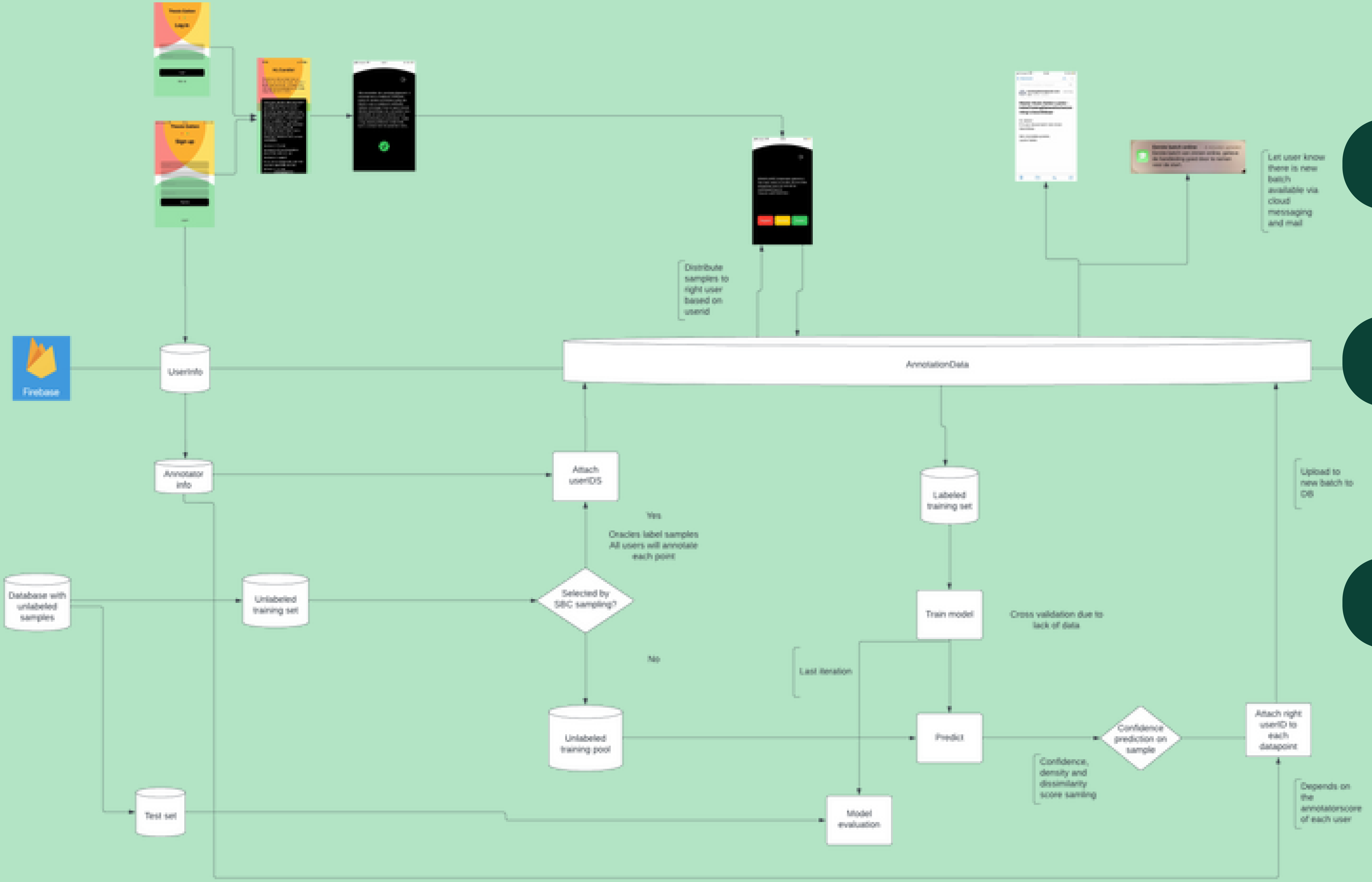
Hey #Overlegcomite, oe zittet just?

Class: Negative

Hallo #Overlegcomite, oa zittet just?

Final proposed solution

GHENT UNIVERSITY



Front end

Database

Back end

Front-end

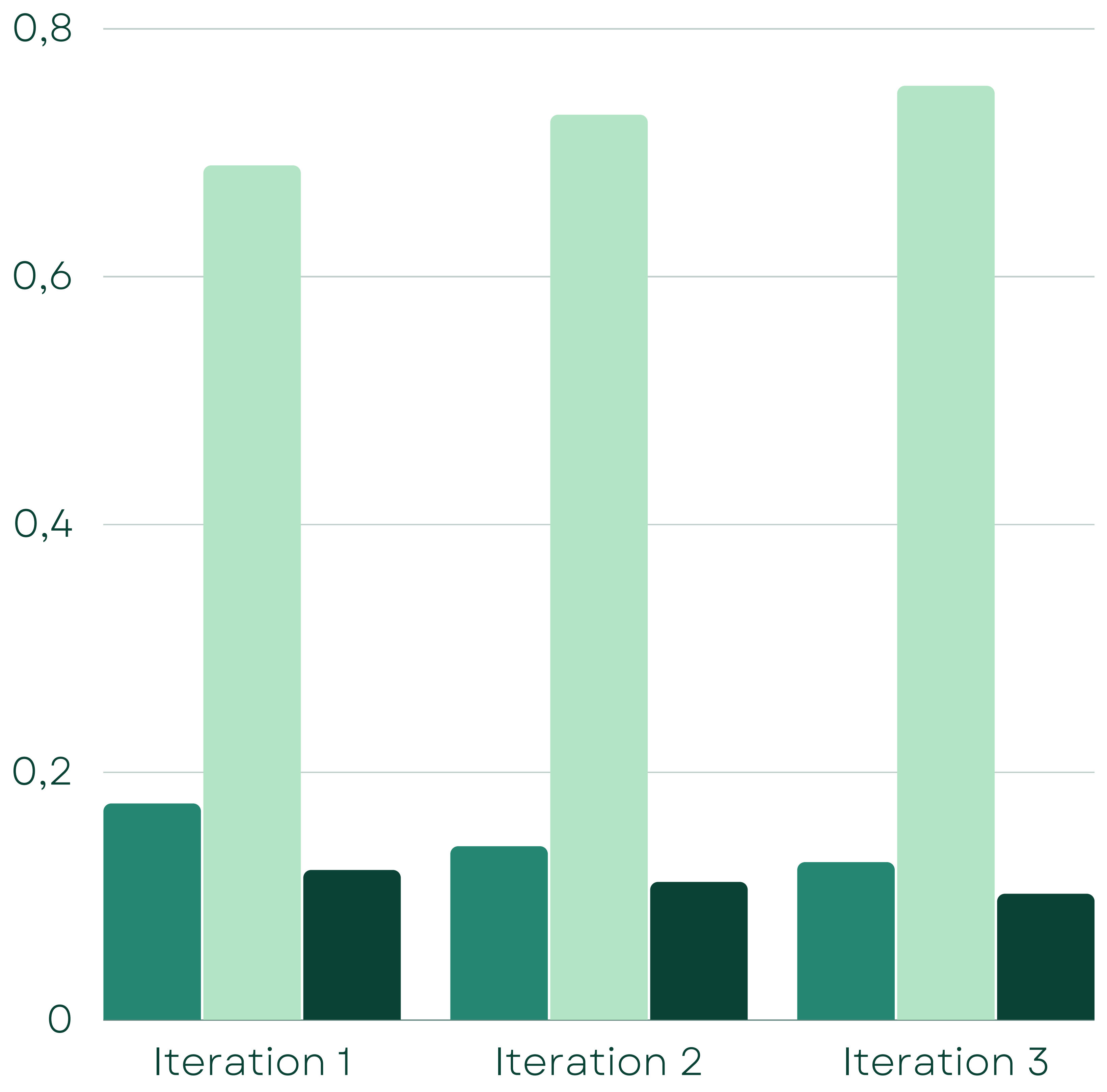
GHENT UNIVERSITY

Encountered challenge – class imbalance

Class imbalance

Class imbalance became worse

Negative effects get enhanced by low amount of data



Big deal? Yes!

GHENT UNIVERSITY

01

Generalization

Very limited examples from minority class, neural network struggles to generalize to other examples

02

Biased decision boundaries

Decision boundaries are biased towards majority class. As we sample based on uncertainty, higher chance of sampling neutral. Reason why it kept increasing!

03

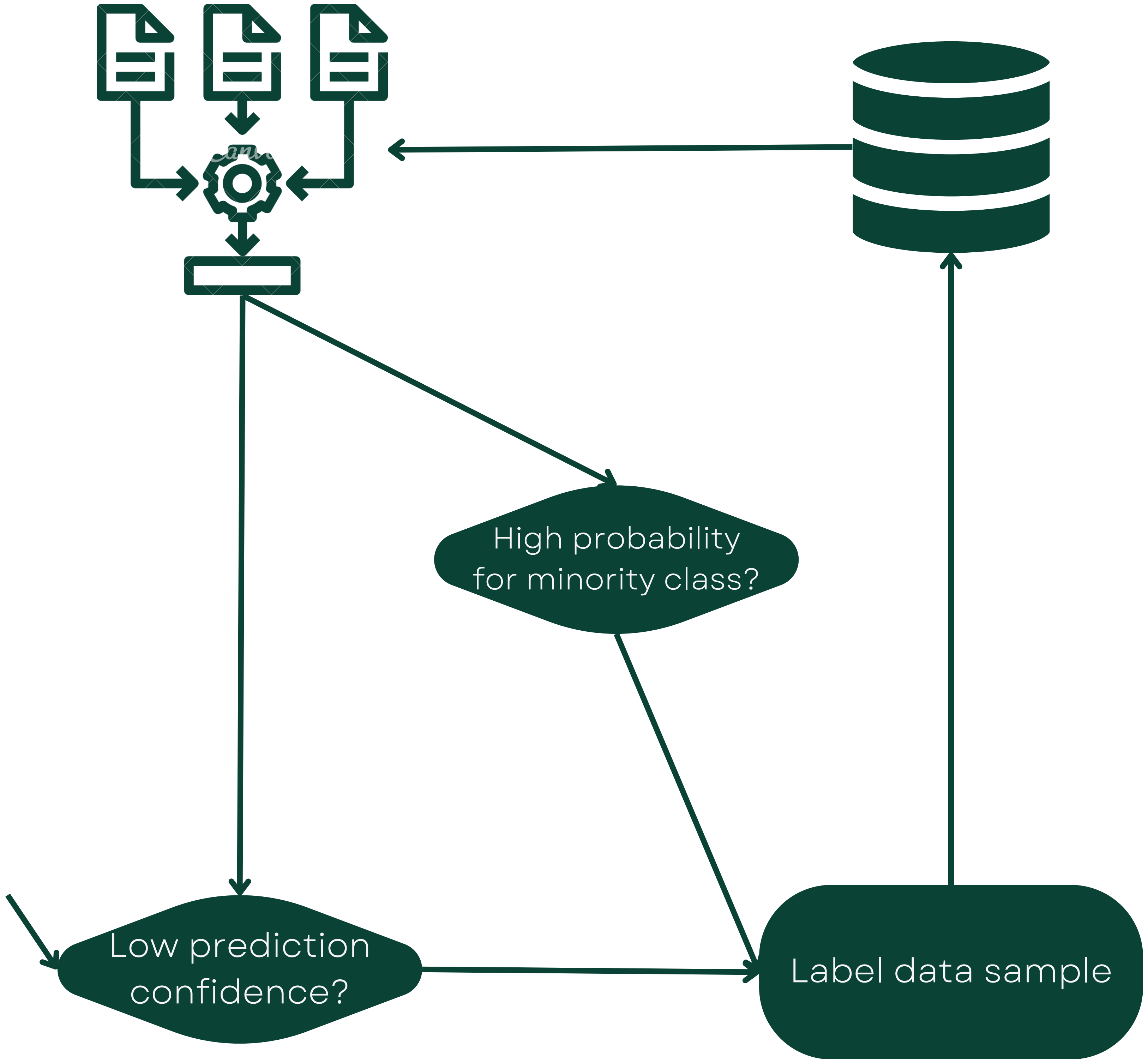
Overfitting majority class

Especially problematic with such a heavy model and low amount of data. Predicted everything neutral.

Proposed solution – part 1

- Use predicted probabilities from model to sample from minority class
- However still limited solution in the beginning as there is a generalization issue
- Still not certain that sampled points will belong to a minority class
- Over- and undersampling not an option. Will generate doubles and overfitting due to lack of data

Source: Aggarwal, U., Popescu, A., & Hudelot, C. (2020). Active Learning for Imbalanced Datasets. https://doi.org/10.1109/wacv45572.2020.9093475



GHENT UNIVERSITY

Proposed solution - part 2 - INS

GHENT UNIVERSITY

Shift decision boundaries towards minority classes

We will sample more of minority classes when sampling uncertainty based

We will sample more from minority class in part 1

Learn more discriminative features for better generalization

Faster convergence and more stable training dynamics

Evaluation

GHENT
UNIVERSITY

Evaluation

	Reference model	Active learning
Size train set	7779	1389 (59 attacks)
Accuracy	82.97%	81.32% (no attacks: 81.24%)
F-1 score	0.594328	0.566336 (no attacks: 0.564)
Negatives/Neutrals/Positives	25.23%/60.79%/13.97%	15.05%/76.46%/8.49%

However...

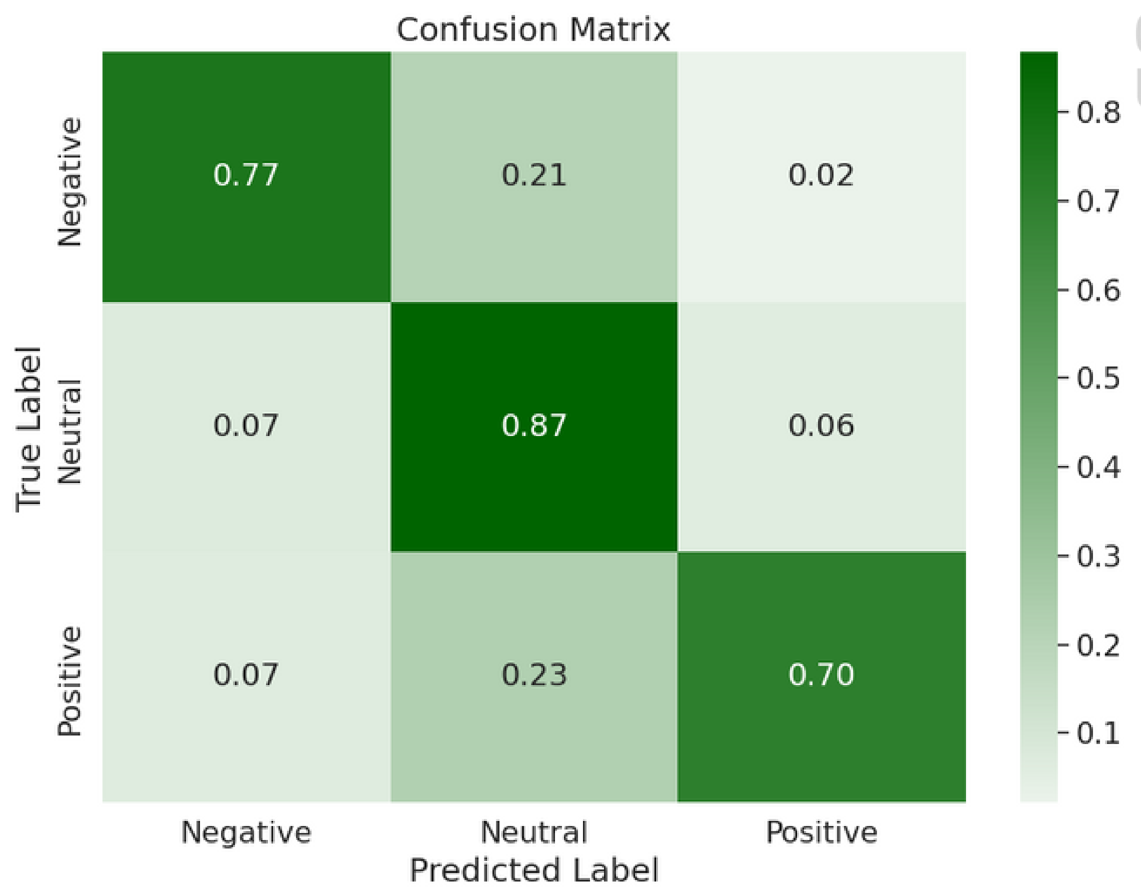
Some context needed

- Probably overestimated;
  Test set:
  composed by two annotators -> not agree -> not included

- Depending on our main goal, we could make some changes to the weight classes of the loss function

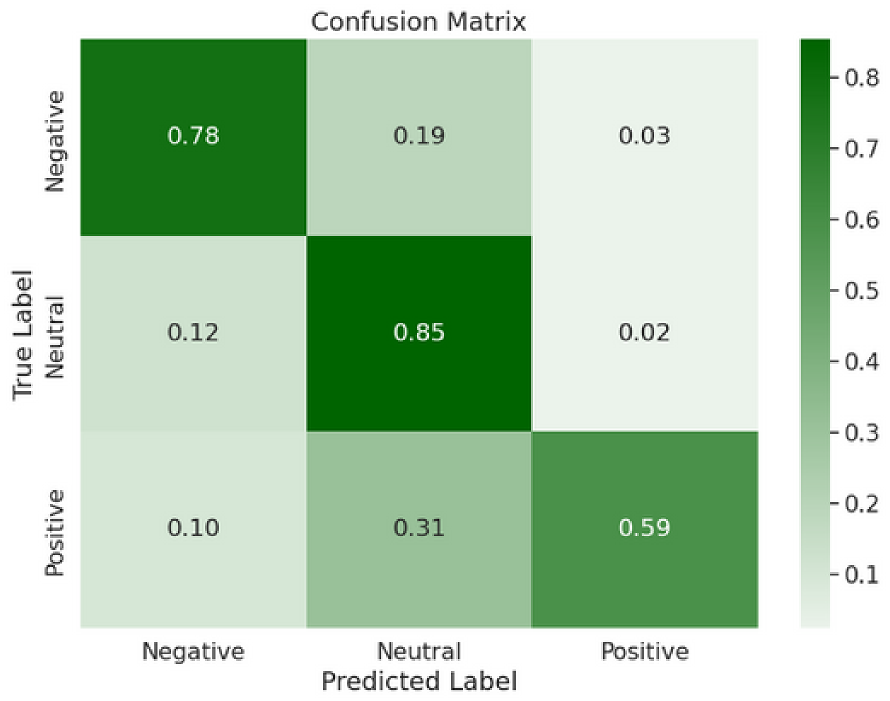
- Still struggles with identifying positives

Reference model

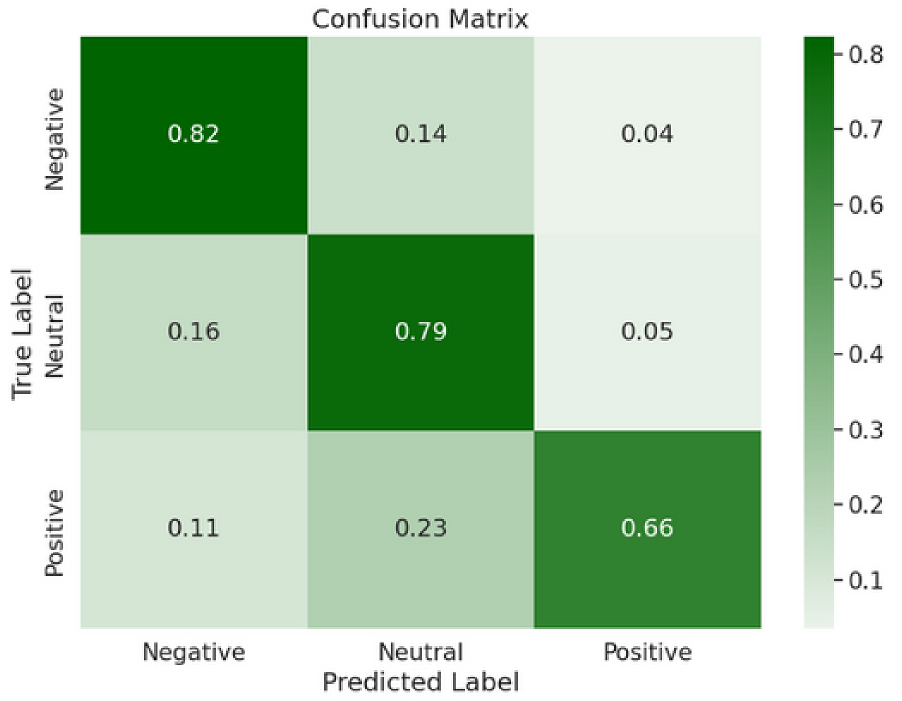


Active learning model

Smaller weights



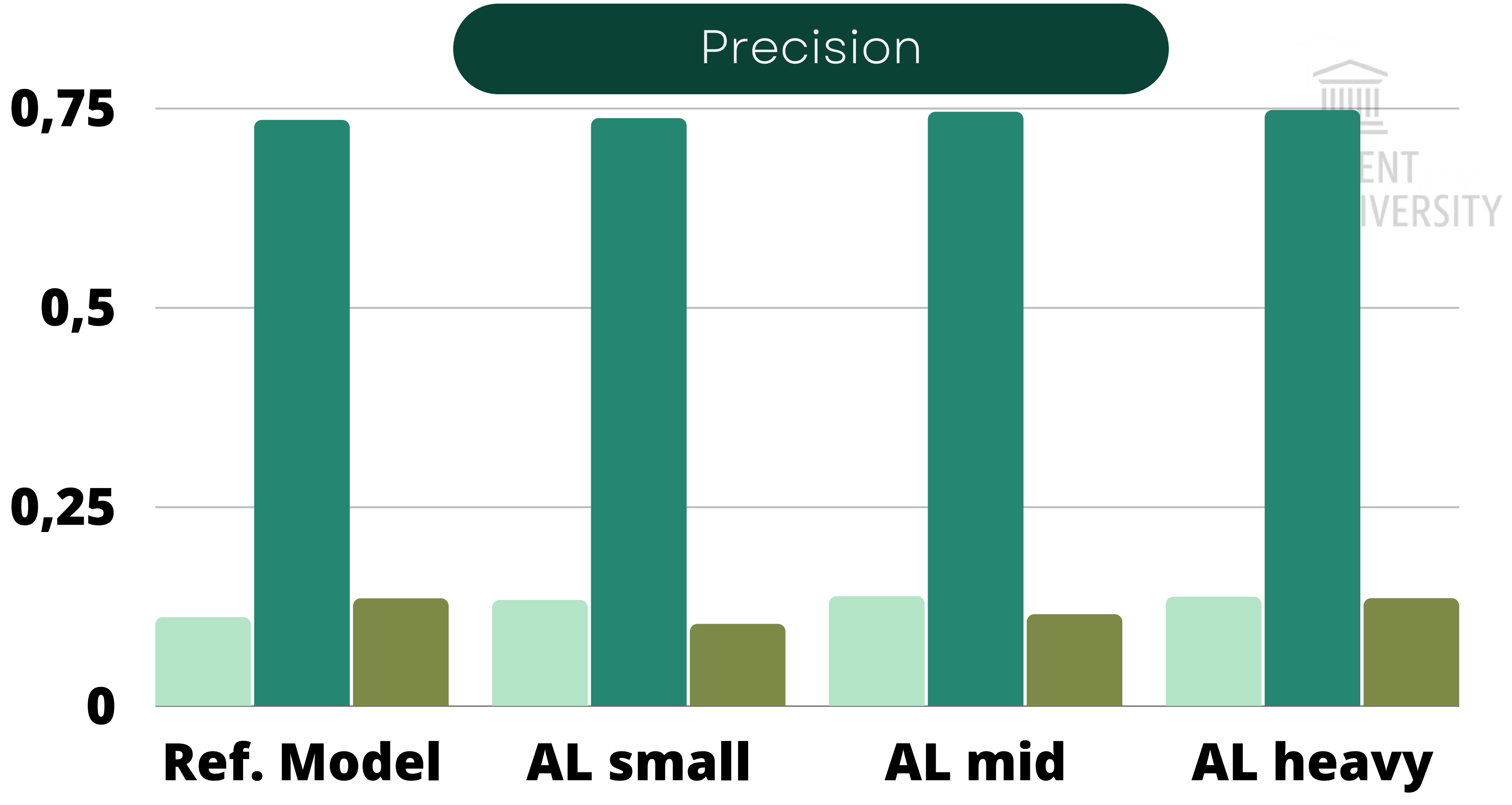
Bigger weights

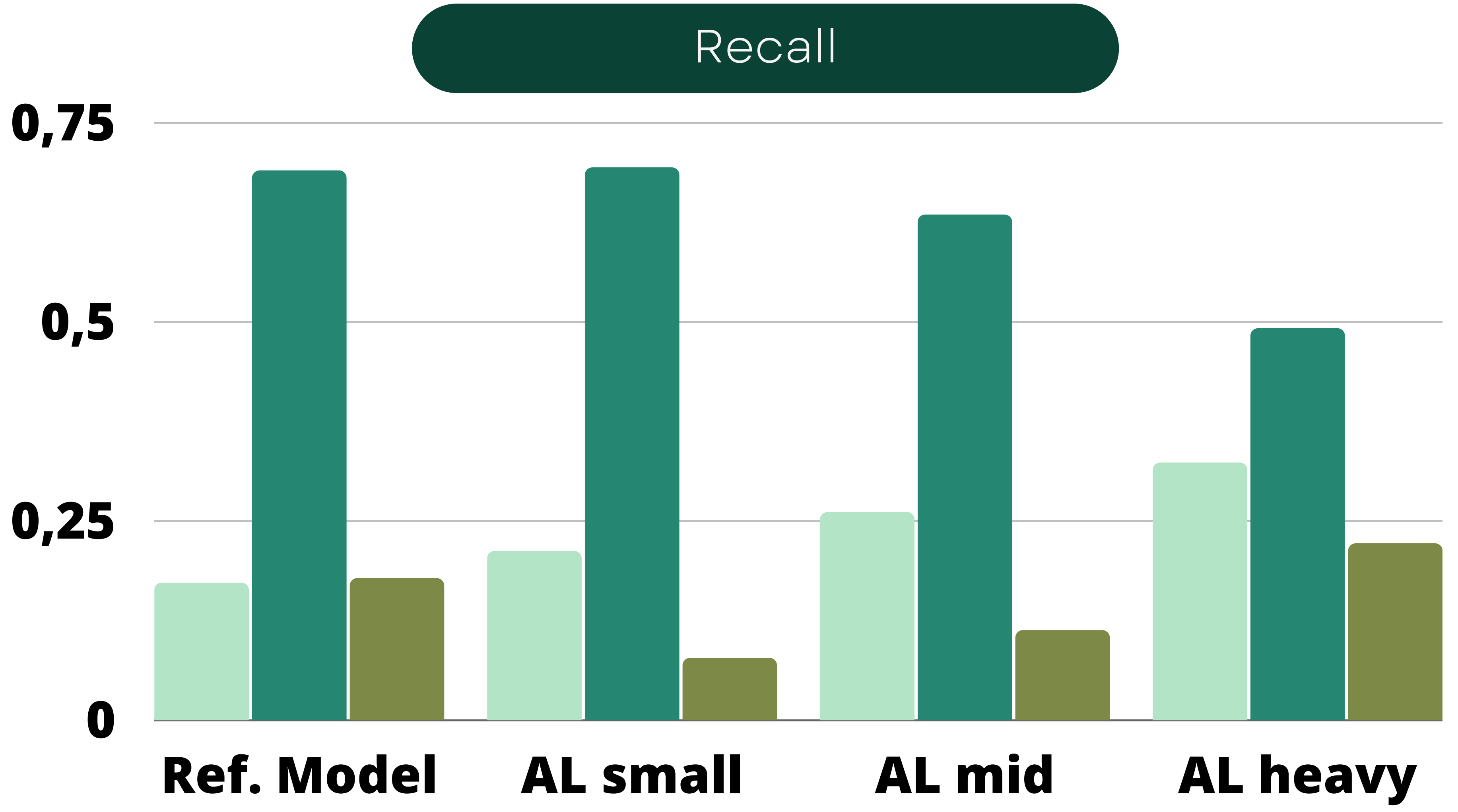


However...

Some context needed

- AL with low class weight differences has highest accuracy
- AL with heavy weight differences is better in recognizing the minority classes
- At cost of mis-classifying neutrals

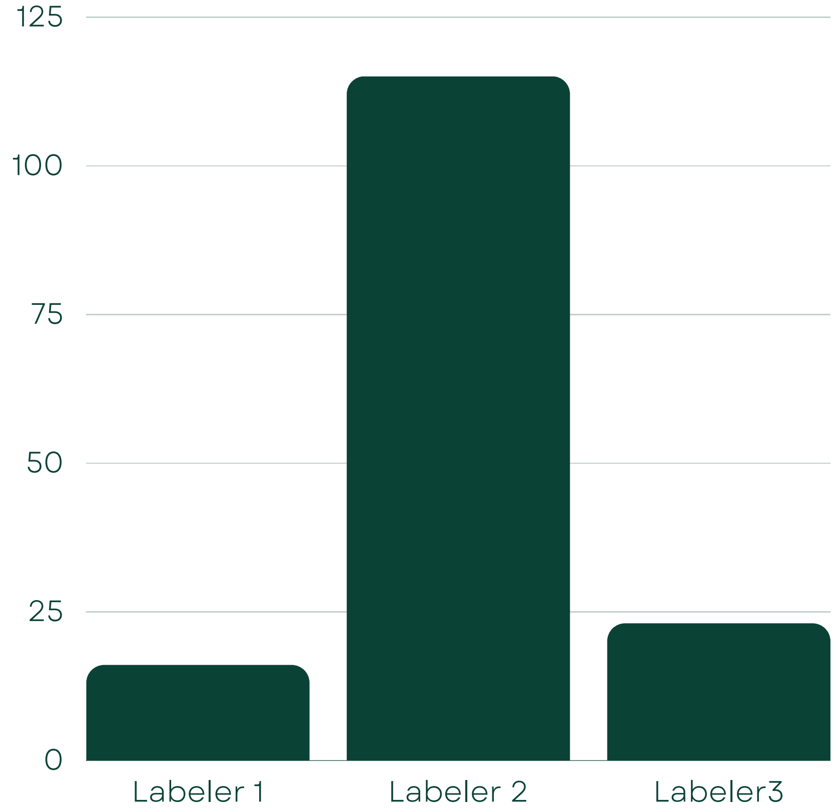




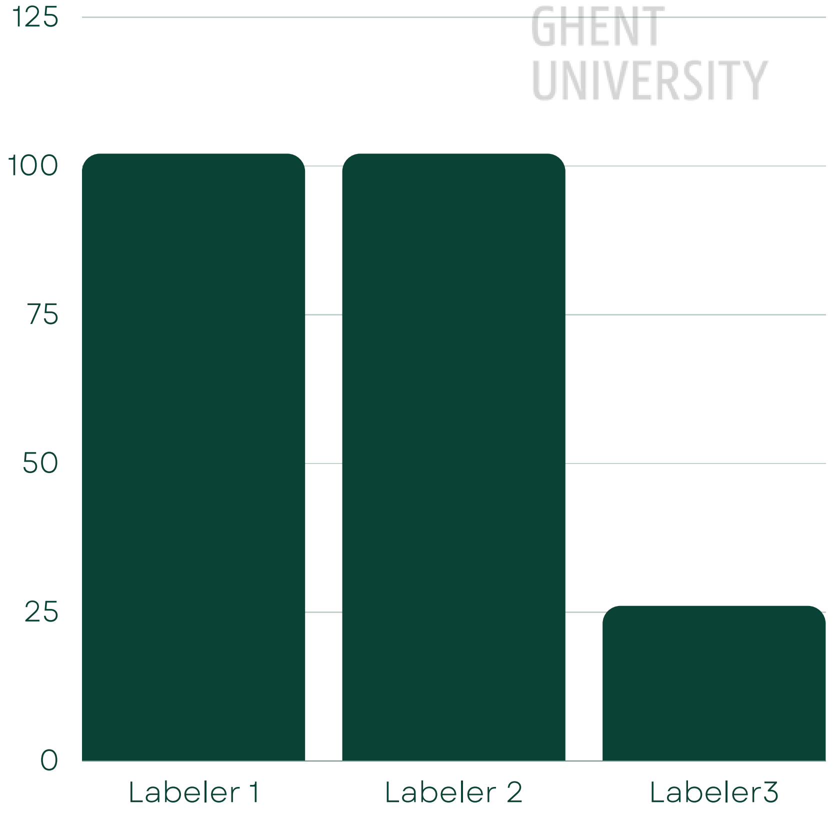
Annotator-quality

Different annotators get different topics

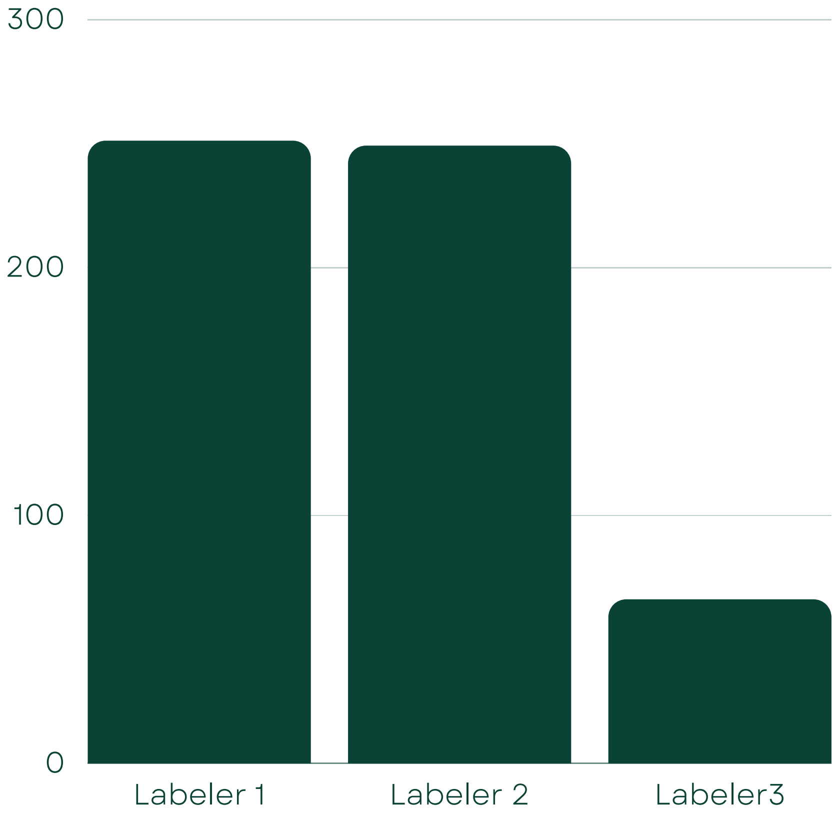
Thema 0



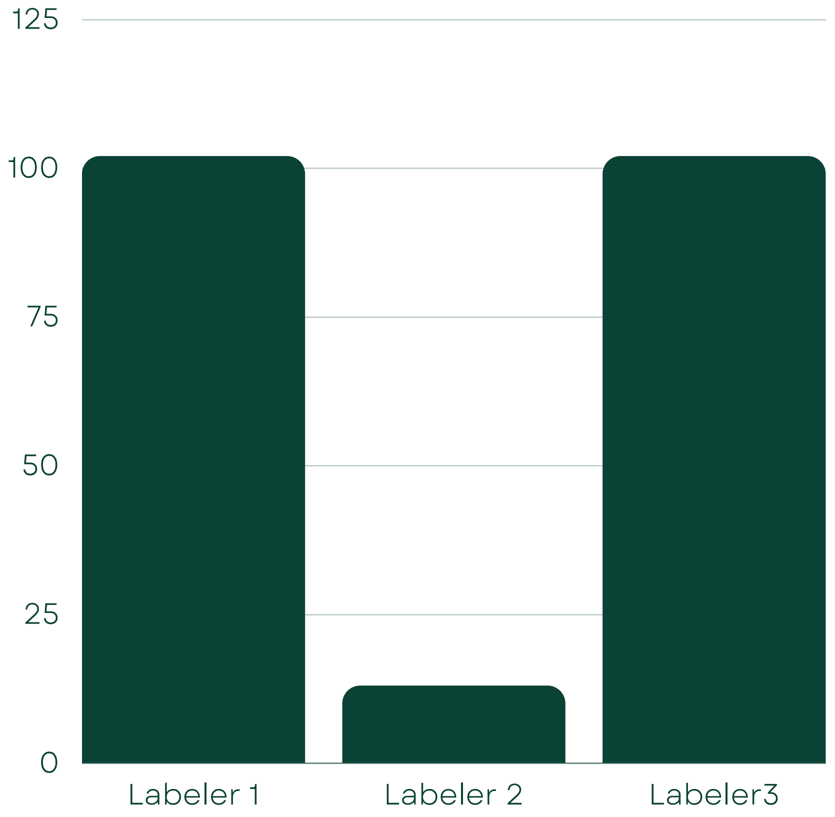
Thema 1



Thema 2



Thema 3



Conclusion

Very interesting for specific (business) cases

Where crowd-sourcing is limited and available labeled date is scarce

Experiment setup was not flawless

We could even further reduce amount of labeled data

Imbalanced data

Active learning with imbalanced data is challenging but not invincible

Hyperparameter optimization might be needed

There ain't no such thing as a free lunch

Shifts work from annotators to computing time, experiment time and setup time

Questions?

GHENT UNIVERSITY