NLP and Semantic Measures for analysing psychological constructs

By Matilda Andersson and Alexandra Antgren

Problem

Research questions

Implementation

Results

Conclusion and further work

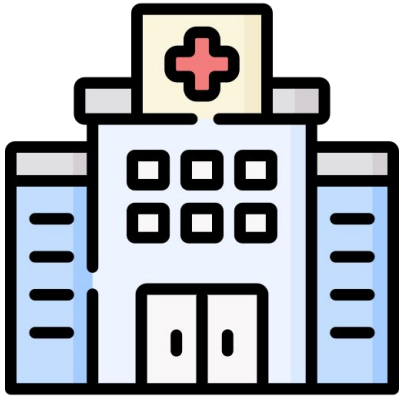
Problem

Research questions

Implementation

Results

Conclusion and further work



Icons from Flaticon

*Rate each of the following statements on a scale of 1 ("not at all typical of me") to 5 ("very typical of me"). Please do not leave any items blank.

	Not at all typical of me 1	2	3	4	Very typical of me 5
1. If I do not have enough time to do everything, I do not worry about it.	○	○	○	○	○

- Current method: numerical rating scales → the dominant method for measuring people's mental state

- Our method: Open-ended responses→ semantic space → explore whether the semantic space correlates with if a person has anxiety/depression or not.

Problem

Research questions

Implementation

Results

Conclusion and further work

1) How well can question-based computational language assessments (QCLAs) correlate with traditional numerical rating scales?

2) Can QCLAs and semantic representations be used for predicting depression/anxiety?

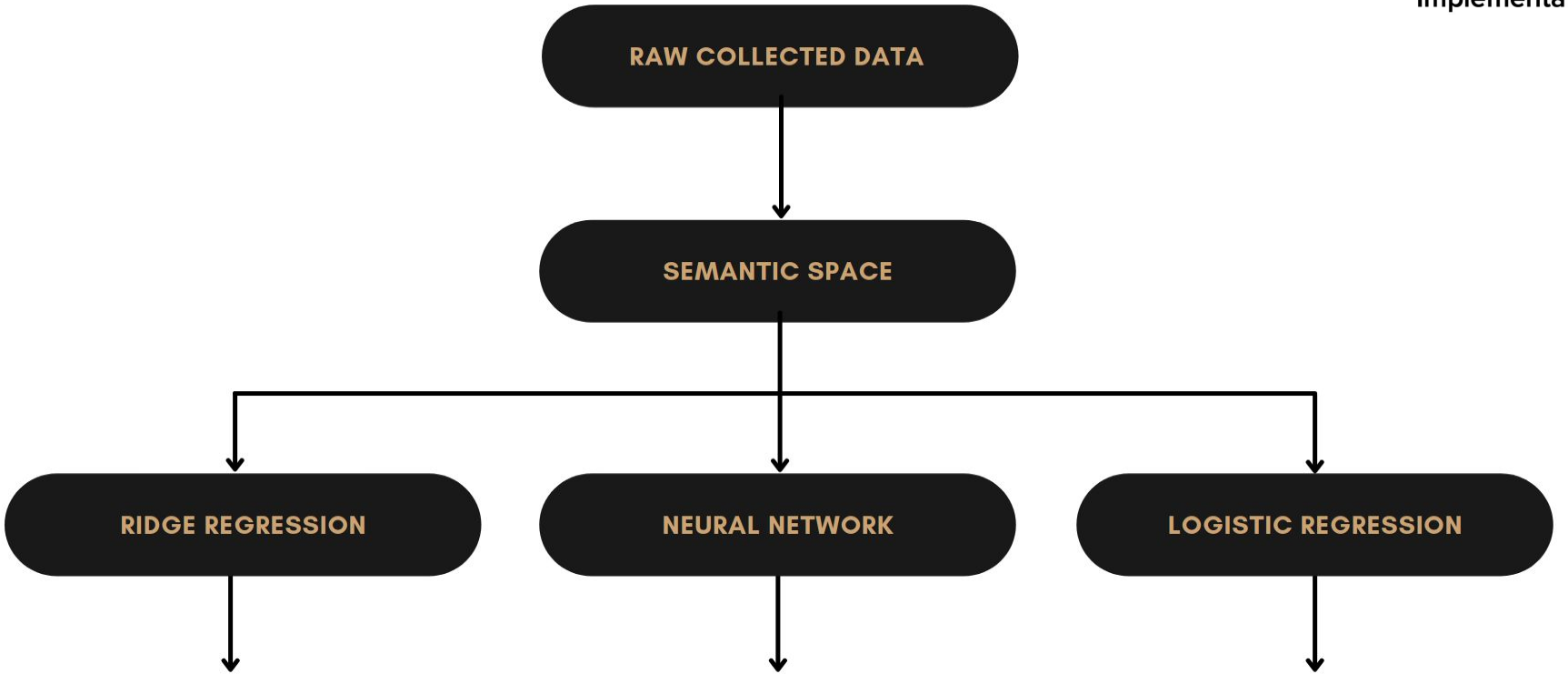
Problem

Research questions

Implementation

Results

Conclusion and further work


RAW COLLECTED DATA
SEMANTIC SPACE
RIDGE REGRESSION
NEURAL NETWORK
LOGISTIC REGRESSION
NUMERICAL VALUE
CORRESPONDING TO
PHQ9 OR GAD7 SCALES
CLASS PREDICTION
ORDINAL CLASSES OF
CONDITION SEVERITY
BINARY PREDICTION
DIAGNOSE DEPRESSION
/ANXIETY OR NOT?

RAW COLLECTED DATA

*Over the last 2 weeks, have you been worried or not?

Write descriptive words relating to those aspects that are most important and meaningful to you.

Write only one descriptive word in each box.

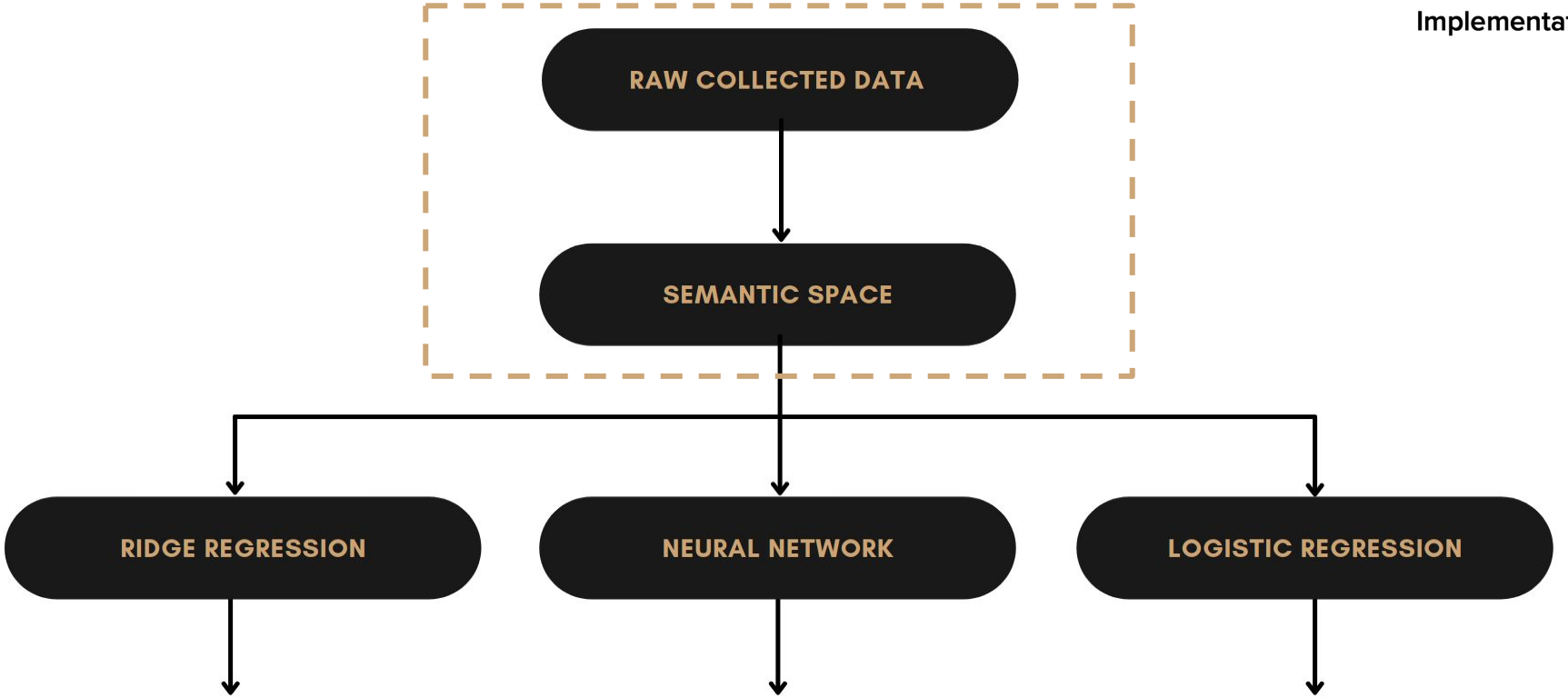
Word 1		Word or phrase 1	
Word 2		Word or phrase 2	
Word 3		Word or phrase 3	
Word 4		Word or phrase 4	
Word 5		Word or phrase 5	

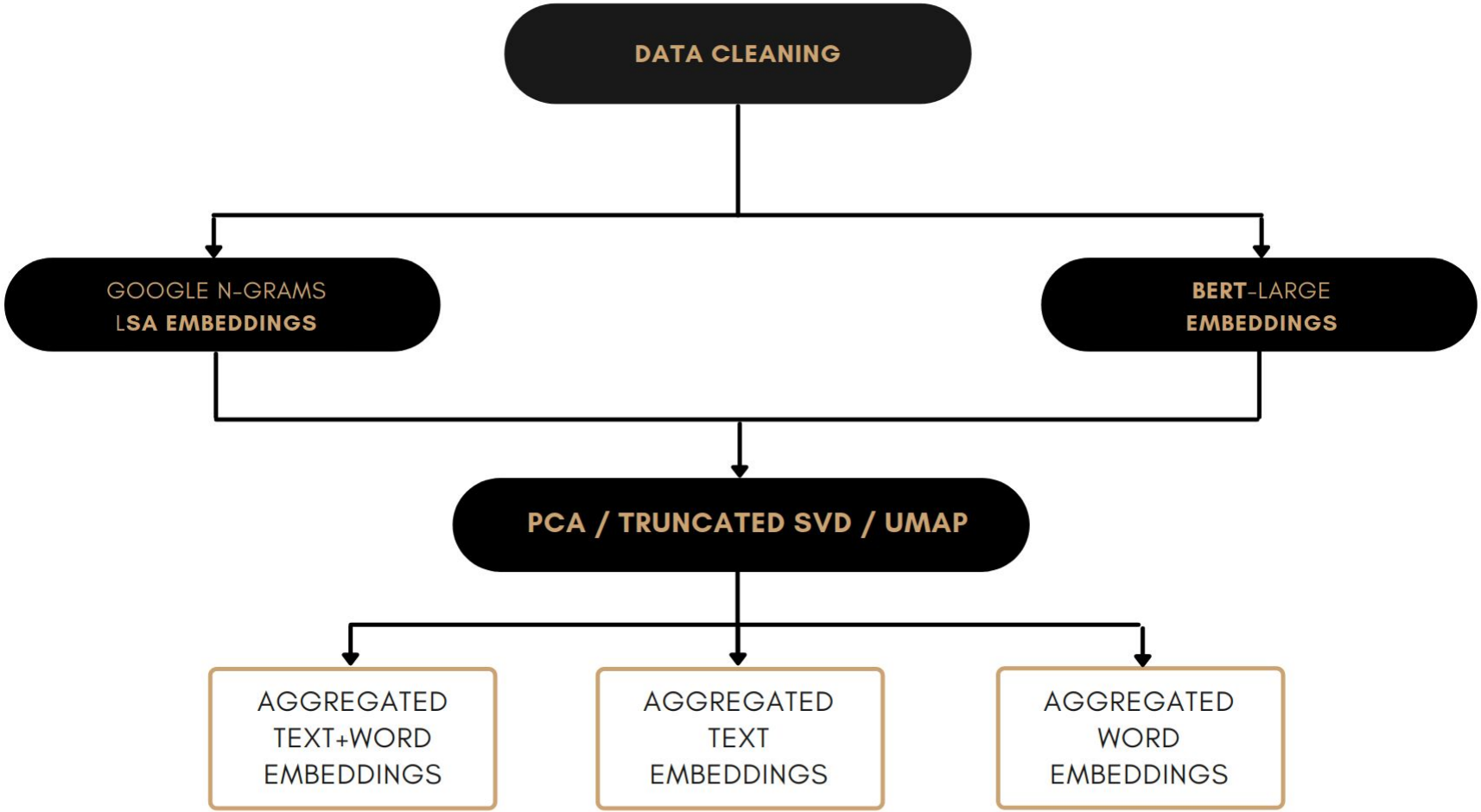
RIDGE REG

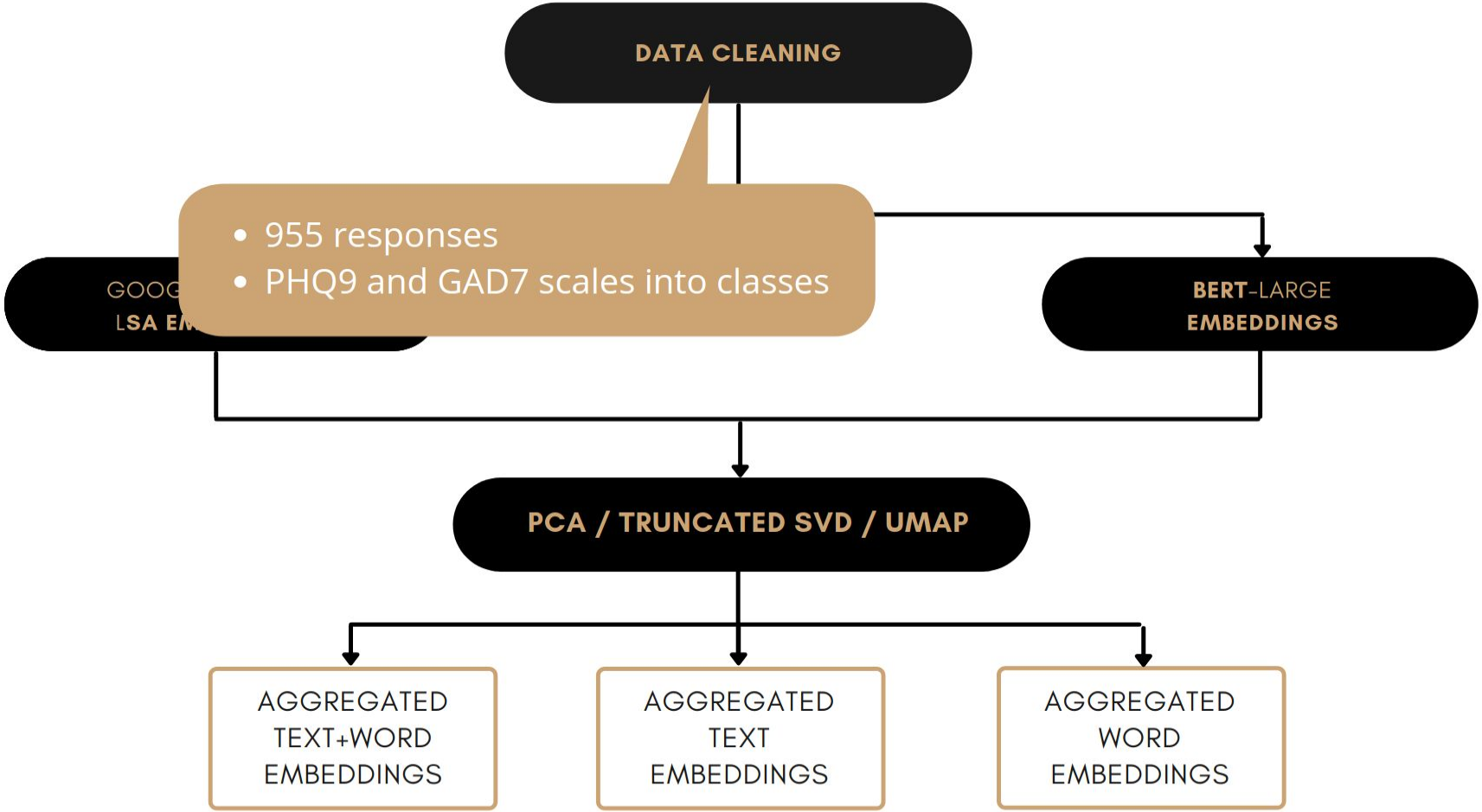
NUMERICA

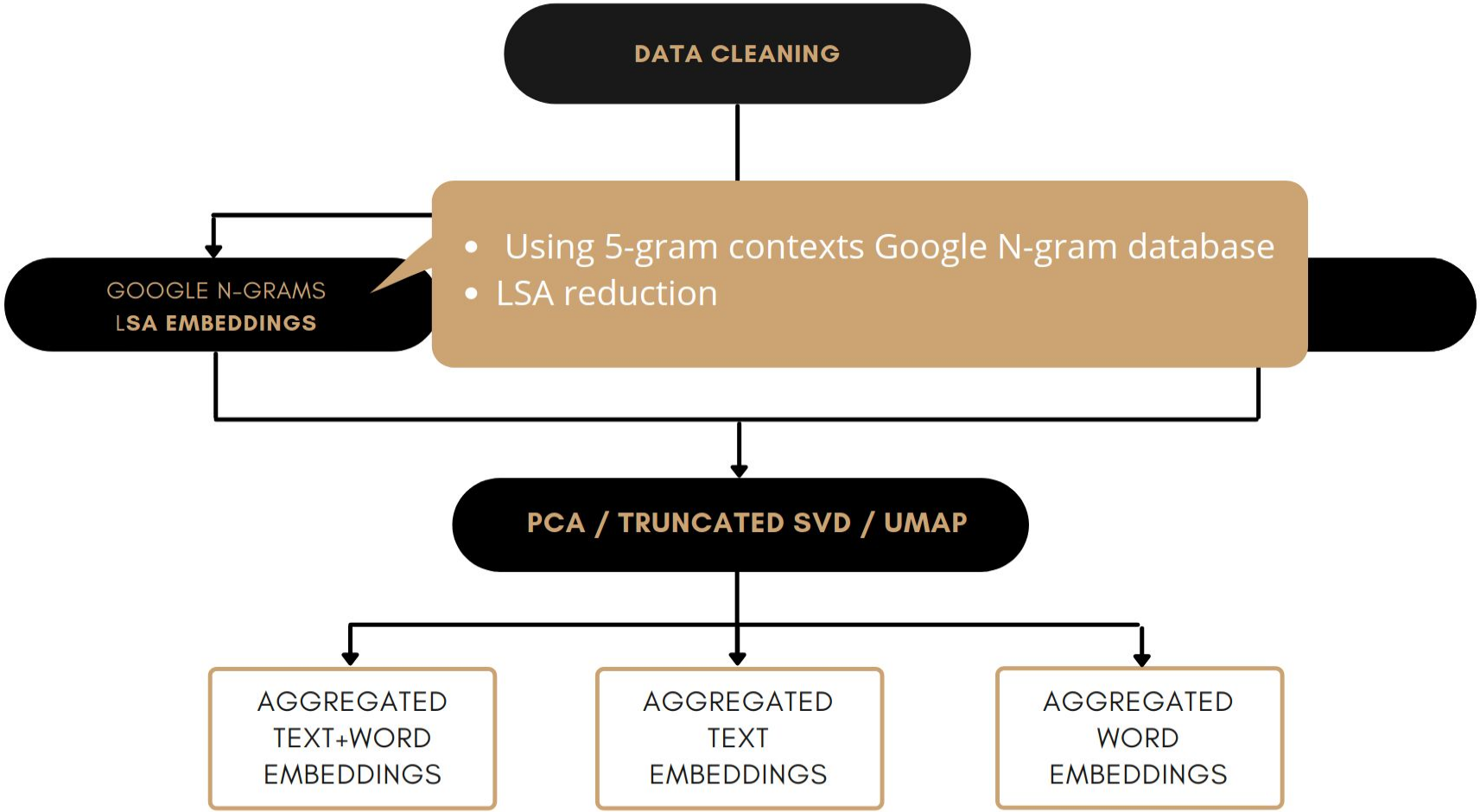
CORRESPO

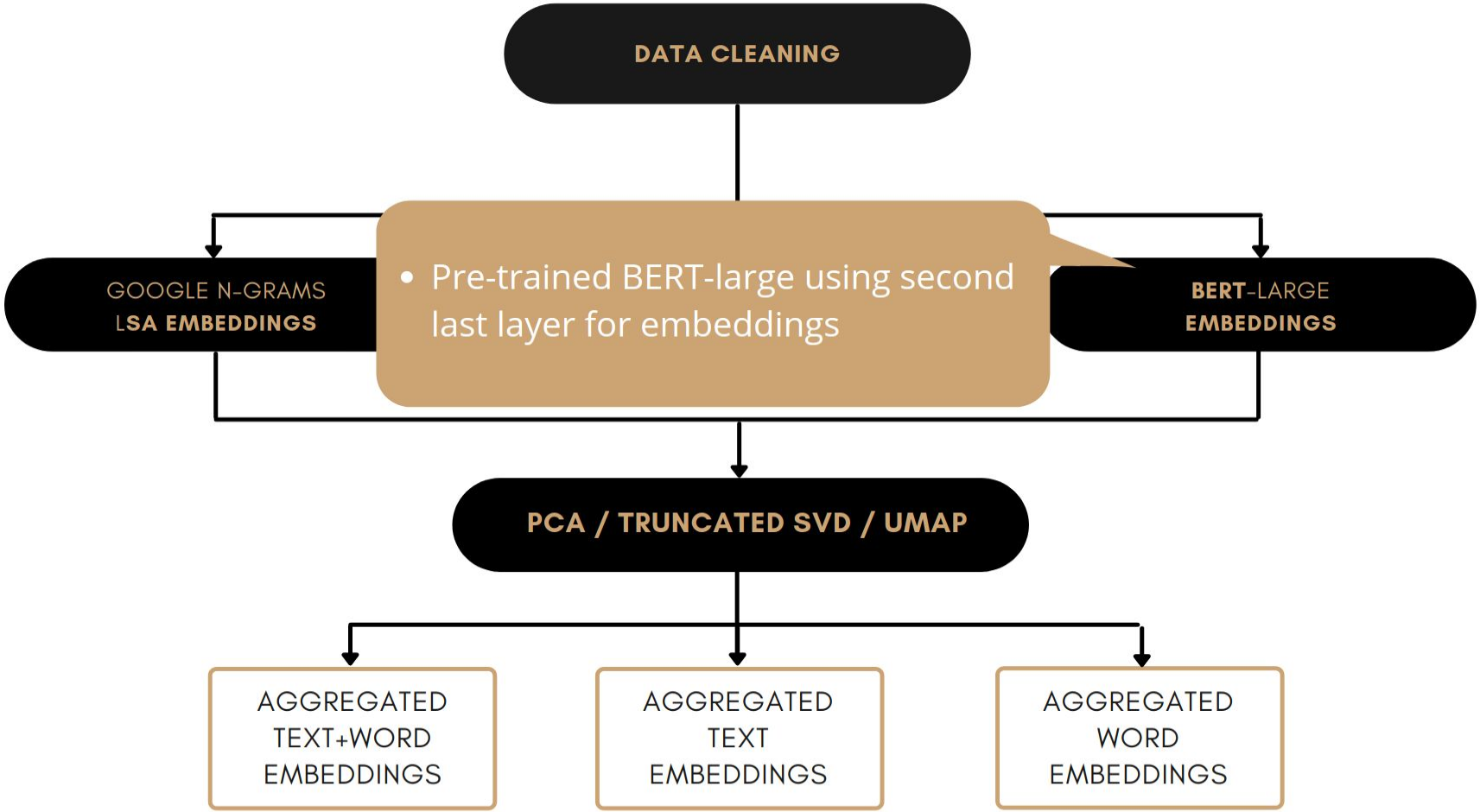
PHQ9 OR GA

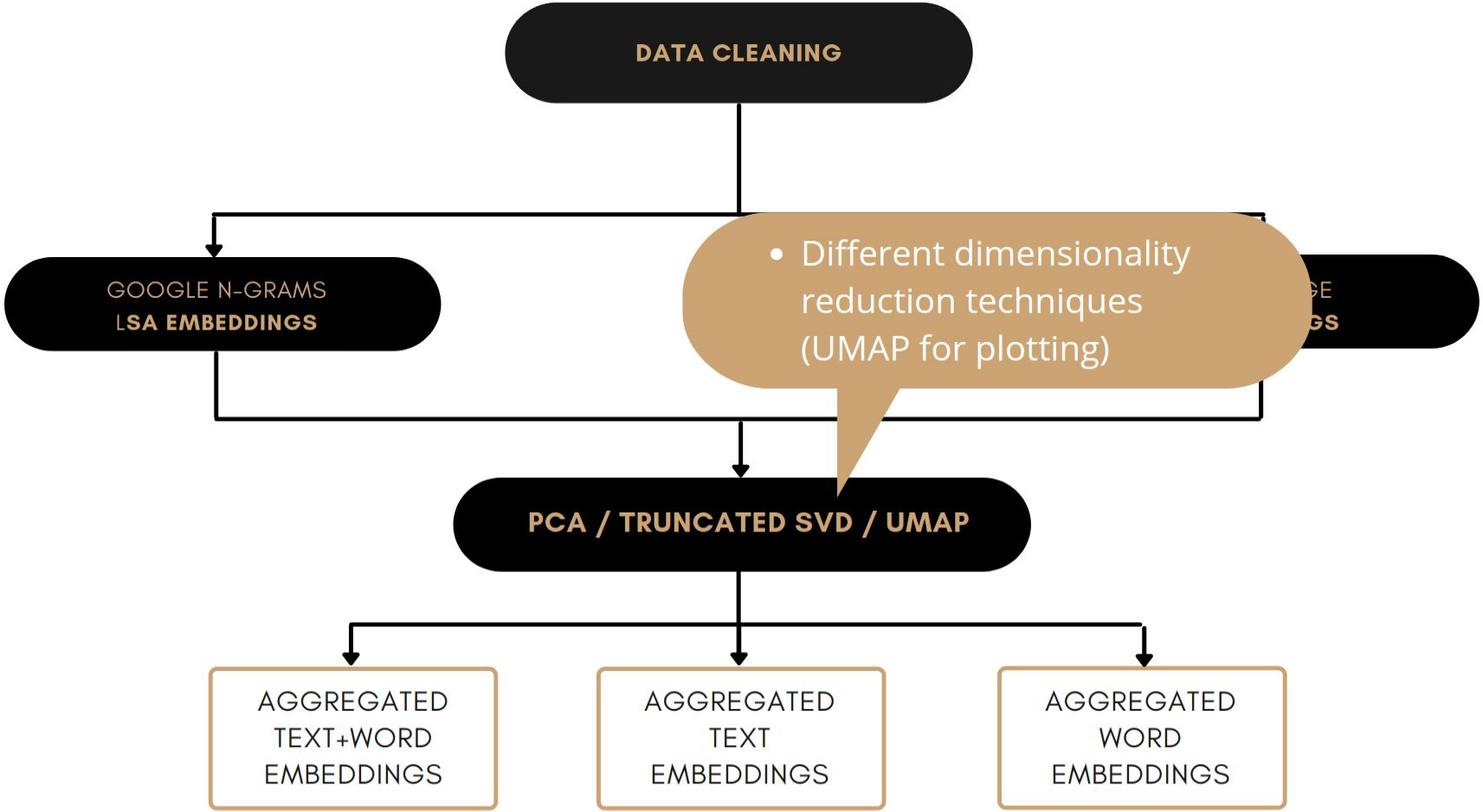

RAW COLLECTED DATA
SEMANTIC SPACE
RIDGE REGRESSION
NEURAL NETWORK
LOGISTIC REGRESSION
NUMERICAL VALUE
CORRESPONDING TO
PHQ9 OR GAD7 SCALES
CLASS PREDICTION
ORDINAL CLASSES OF
CONDITION SEVERITY
BINARY PREDICTION
DIAGNOSE DEPRESSION
/ANXIETY OR NOT?

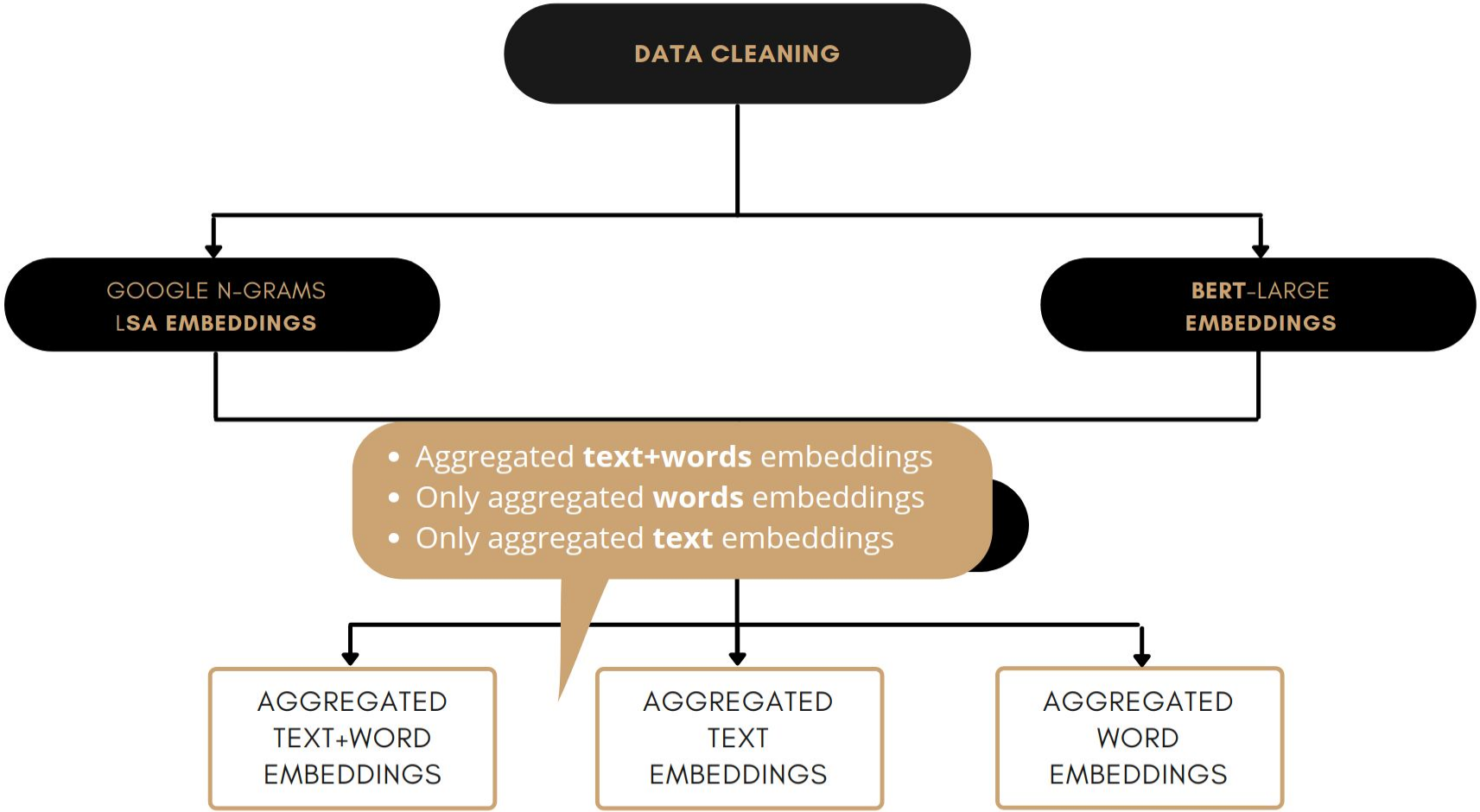

DATA CLEANING
GOOGLE N-GRAMS
LSA EMBEDDINGS
BERT-LARGE
EMBEDDINGS
PCA / TRUNCATED SVD / UMAP
AGGREGATED
TEXT+WORD
EMBEDDINGS
AGGREGATED
TEXT
EMBEDDINGS
AGGREGATED
WORD
EMBEDDINGS

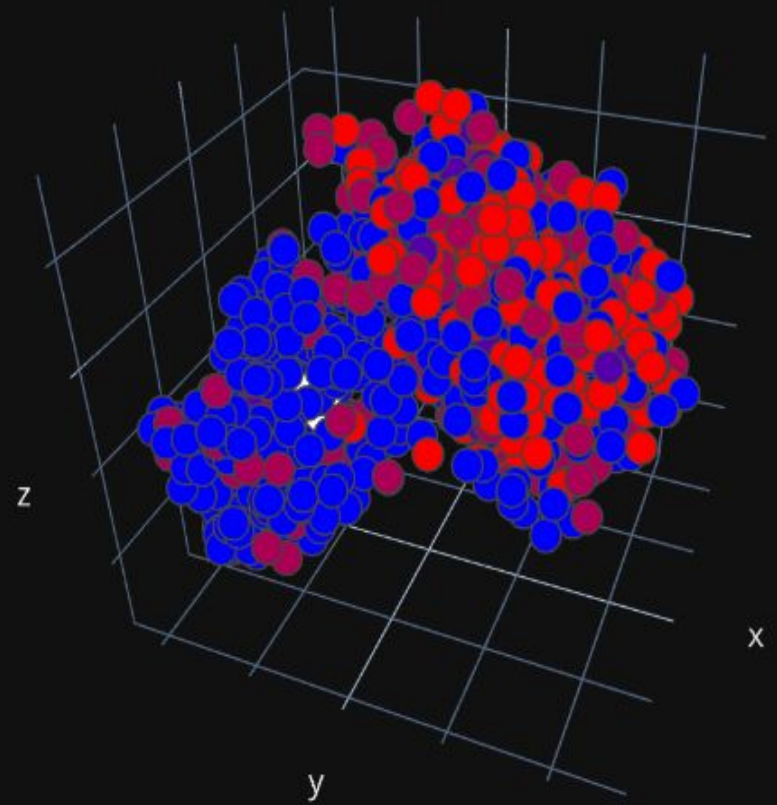

DATA CLEANING
• 955 responses
• PHQ9 and GAD7 scales into classes
GOOG
LSA EM
BERT-LARGE
EMBEDDINGS
PCA / TRUNCATED SVD / UMAP
AGGREGATED
TEXT+WORD
EMBEDDINGS
AGGREGATED
TEXT
EMBEDDINGS
AGGREGATED
WORD
EMBEDDINGS


DATA CLEANING
GOOGLE N-GRAMS
LSA EMBEDDINGS
• Using 5-gram contexts Google N-gram database
• LSA reduction
PCA / TRUNCATED SVD / UMAP
AGGREGATED TEXT+WORD EMBEDDINGS
AGGREGATED TEXT EMBEDDINGS
AGGREGATED WORD EMBEDDINGS


DATA CLEANING
GOOGLE N-GRAMS LSA EMBEDDINGS
• Pre-trained BERT-large using second last layer for embeddings
BERT-LARGE EMBEDDINGS
PCA / TRUNCATED SVD / UMAP
AGGREGATED TEXT+WORD EMBEDDINGS
AGGREGATED TEXT EMBEDDINGS
AGGREGATED WORD EMBEDDINGS


DATA CLEANING
GOOGLE N-GRAMS
LSA EMBEDDINGS
• Different dimensionality reduction techniques (UMAP for plotting)
PCA / TRUNCATED SVD / UMAP
AGGREGATED TEXT+WORD EMBEDDINGS
AGGREGATED TEXT EMBEDDINGS
AGGREGATED WORD EMBEDDINGS


DATA CLEANING
GOOGLE N-GRAMS
LSA EMBEDDINGS
BERT-LARGE
EMBEDDINGS
• Aggregated text+words embeddings
• Only aggregated words embeddings
• Only aggregated text embeddings
AGGREGATED
TEXT+WORD
EMBEDDINGS
AGGREGATED
TEXT
EMBEDDINGS
AGGREGATED
WORD
EMBEDDINGS


z
y
x


3
2.5
2
1.5
1
0.5
0

Problem

Research questions

Implementation

Results

Conclusion and further work

1) How well can QCLAs correlate with traditional numerical rating scales?

- **Spearman's correlations** measures the strength and direction of association between two variables
- Value between -1 and +1

	Depression	Anxiety
Baseline Dummy Classifier	0.102	-0.020
Ridge Regression	0.717	0.648
MLP Neural Network	0.642	0.527

2) Predicting numerical value on PHQ-9 or GAD-7 scales

Depression scale predicted vs. self reported		
	Mean Abs Error	R^2
Baseline	1.232	0.0
Words + Text	0.764	0.528
Words	0.813	0.492
Text	0.838	0.432

Anxiety scale predicted vs. self reported		
	Mean Abs Error	R^2
Baseline	0.994	0.0
Words + Text	0.676	0.420
Words	0.695	0.403
Text	0.752	0.300

2) Predicting **depression/anxiety** severity class

Depression	Precision	Recall	F1-score
Accuracy			0.39
Macro average	0.52	0.39	0.34
Weighted average	0.51	0.39	0.35

Anxiety	Precision	Recall	F1-score
Accuracy			0.42
Macro average	0.45	0.40	0.35
Weighted average	0.46	0.42	0.36

2) Predicting **depression, anxiety** diagnose or not?

Logistic Reg.	No depression diagnose	Depression diagnose	Accuracy	Macro avg.	Weighted avg.
Baseline - F1-score	0.65	0.29	0.53	0.47	0.53
Our model - F1-score	0.77	0.56	0.70	0.66	0.70

Logistic Reg.	No anxiety diagnose	Anxiety diagnose	Accuracy	Macro avg.	Weighted avg.
Baseline - F1-score	0.51	0.49	0.50	0.50	0.50
Our model - F1-score	0.64	0.70	0.68	0.67	0.67

Problem

Research questions

Implementation

Results

Conclusion and further work

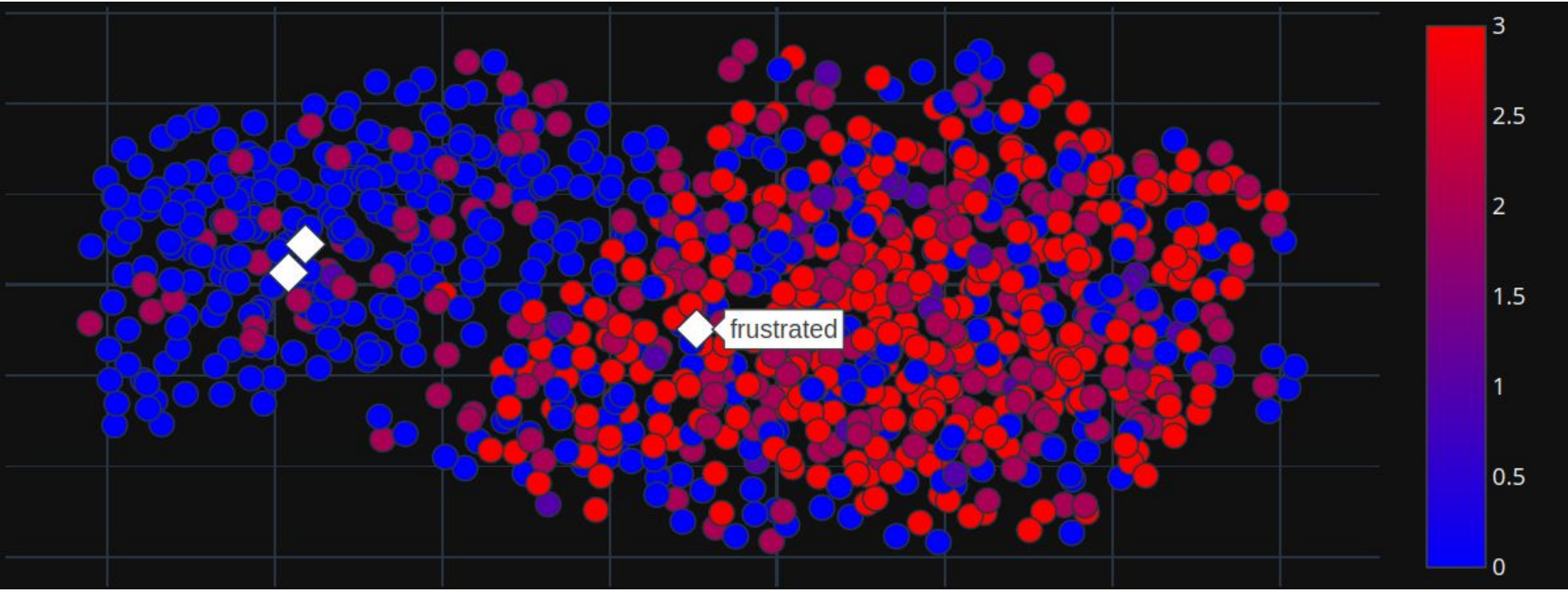
Conclusions and future work

- Results show that semantic measures correlate well with numerical rating scales
- However, the reliability of a self reported scale provides limits on how well they can be predicted
- Future work could focus on a broader range of psychological constructs where evaluation is not based on self reported numerical scales but objective measures like clinical interviews

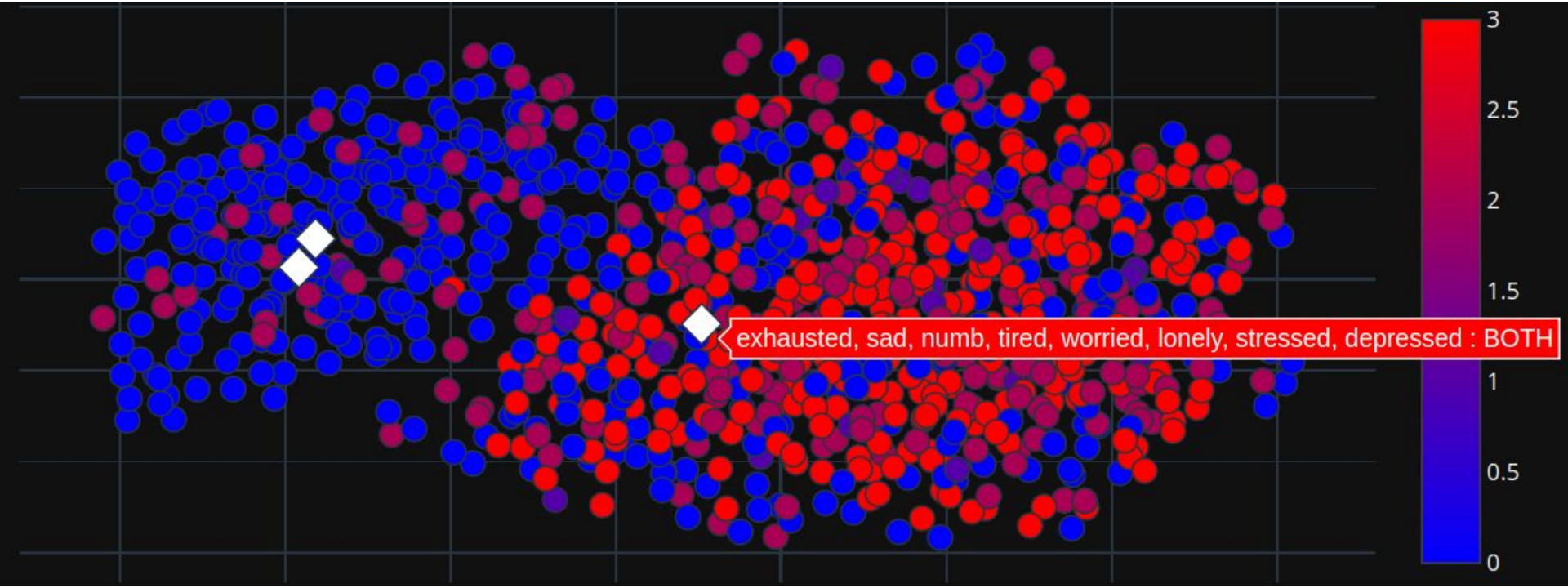
Thank you for listening!

Questions?

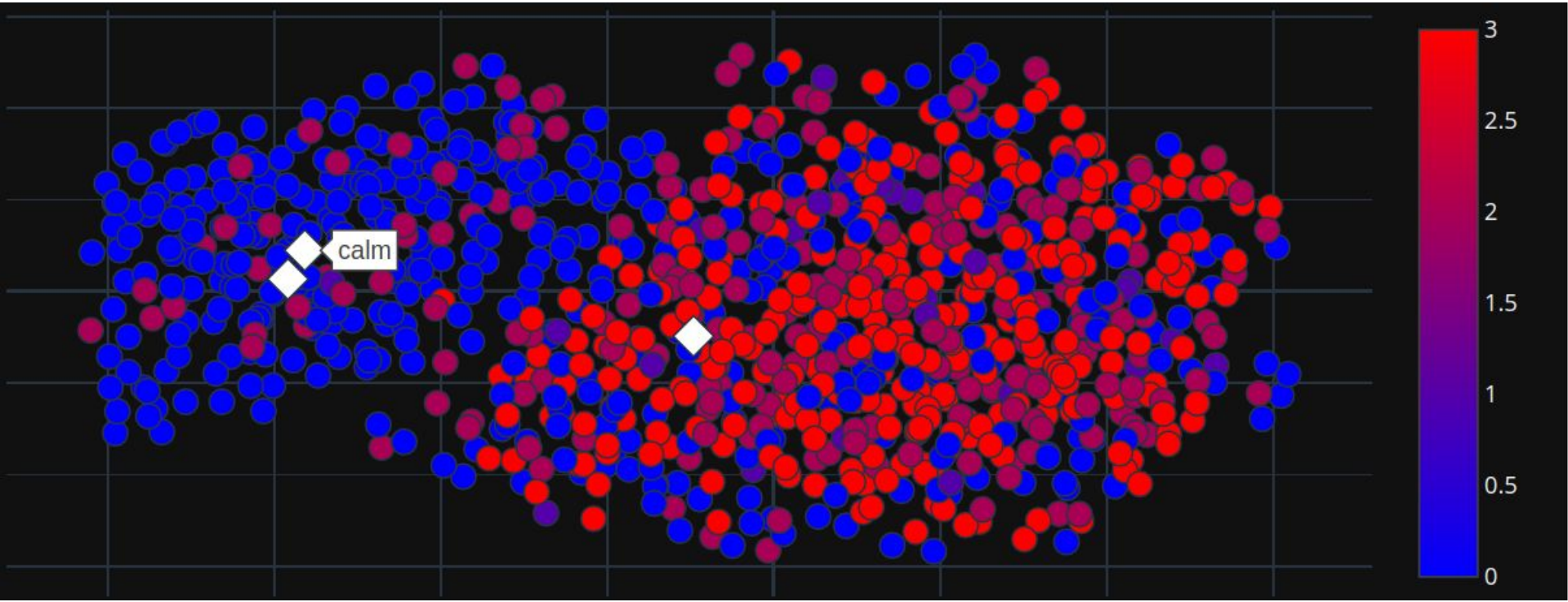
3) Visualizations



3) Visualizations



3) Visualizations



3) Visualizations