Automatic Psychological Text Analysis
Automatic schema classification for Schema-Focused Therapy using k-Nearest Neighbour (kNN)

Budi Han (B.S.Han@student.tudelft.nl)
Professor: Willem-Paul Brinkman
Supervisors: Merijn Bruijnes, Hayley Hung

TUDelft

1. Background

- **Schema-based Therapy** for assessment of **Personality disorders**[1]
- **Short Schema Mode Index** (SMI) for schemas = [*"Vulnerable", "Angry", "Impulsive", "Happy", "Detached", "Punishing", "Healthy"*][1]
- **Automation** realised by Allaart in form of a **Conversational Agent (CA)**[2]
- **Dataset** comprising of stories and SMI questionnaire[2]
- CA allows for multiple **Ecological Momentary assessments** (EMA)
- **Automation** in Schema-Based Therapy will make the treatment more **socially available.**
- **CA lacks proper text analysis and classification algorithm**[2]

2. Research Question

How well can a schema be automatically classified from a text using KNN?

1. **Pre-processing**
2. **Optimal kNN algorithm**
3. **Comparison with SVM and RNN**

3. Method
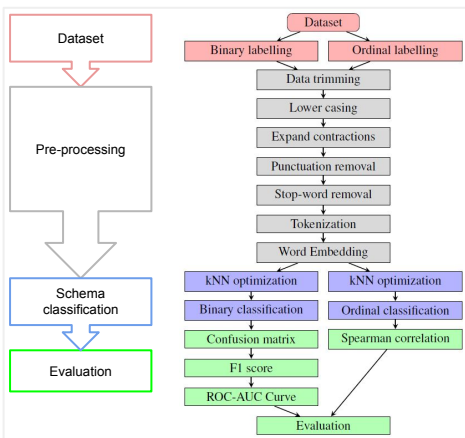
4. Results

Binary classification

Confusion matrix and accuracy	Classification report and f1-score	ROC-AUC curve
- Mean accuracy 71% (57% - 92%) - Healthy predicted true 288 / 289 - Impulsive rather high accuracy	- Mean f1-score 0.45 (0.13 - 0.96) - Healthy **high recall** and precision - Impulsive lowest f1-score - **Contrast** between majority and minority classes	- Micro-average AUC = 0.70 - Macro-average AUC = 0.53 - **Micro > macro** AUC

Table 2: Overview confusion matrices

	TP	TN	FP	FN	Accuracy
vulnerable	26	163	31	69	0.65
angry	42	122	58	67	0.57
impulsive	5	218	9	57	0.77
happy	181	17	62	29	0.69
detached	28	155	38	68	0.63
punishing	10	206	21	52	0.75
healthy	266	0	22	1	0.92
Mean					0.71

Table 3: Classification report binary kNN

schema	precision	recall	f1-score	support
vulnerable	0.46	0.27	0.34	95
angry	0.42	0.39	0.40	109
impulsive	0.36	0.08	0.13	62
happy	0.74	0.86	0.80	210
detached	0.42	0.29	0.35	96
punishing	0.32	0.16	0.22	62
healthy	0.92	1.00	0.96	267
micro avg	0.70	0.62	0.66	901
macro avg	0.52	0.44	0.46	901
weighted avg	0.64	0.62	0.62	901
samples avg	0.74	0.69	0.68	901
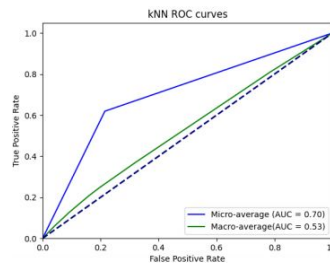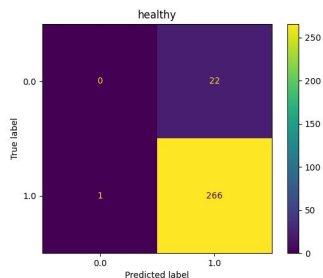
Figure 2: Micro/Macro averaged ROC Curve kNN

Ordinal classification

Spearman correlation

- Low mean 0.09
- No association between predicted and actual values

Table 4: Spearman correlation coefficient per Schema

schema	Spearman
vulnerable	0.13
angry	0.08
impulsive	0.12
happy	0.06
detached	0.08
punishing	0.09
healthy	0.06
mean	0.09

5. Comparison SVM and RNN

Table 6: Overview classifier comparison averaged AUC over all schemas

AUC	SVM	kNN	RNN
micro avg	**0.78**	0.70	0.69
macro avg	0.51	0.53	**0.54**

Table 5: Classifier comparison f1-score

schema	SVM	kNN	RNN
vulnerable	0.27	0.34	0.38
angry	0.38	0.40	0.48
impulsive	0.07	0.13	0.17
happy	0.75	0.80	0.77
detached	0.18	0.35	0.36
punishing	0.20	0.22	0.34
healthy	0.93	0.96	0.95
schema avg	0.39	0.45	0.50
micro avg	0.63	0.66	0.66
macro avg	0.40	0.46	0.49
weighted avg	0.57	0.62	0.63
samples avg	0.66	0.68	0.66

6. Discussion and limitations

- **No reliable foundation for evaluation**
  - **Imbalanced dataset**
  - **Incorrect labeling**

7. Conclusion and future work

- **Conclusion**
  - Satisfactory in binary classification
  - Inadequate in ordinal classification
  - RNN best
  - **Unreliable conclusion** due to limitations
- **Future work**
  - Field experiment
  - More stores, more reliable data
  - Connection between stories and SMI questionnaire

8. References

[1] J. Lobbestael, M. v. Vreeswijk, P. Spinhoven, E. Schouten, and A. Arntz, *"Reliability and Validity of the Short Schema Mode Inventory (SMI),"* Behavioural and Cognitive Psychotherapy, vol. 38, no. 4, pp. 437– 458, 2010. Publisher: Cambridge University Press.

[2] D. Allaart, *"Schema mode assessment through a conversational agent,"* Delft University of Technology, 2021.