Measuring the Severity of the Signs of Eating Disorders Using Similarity-Based Models

Seyed Habib Hosseini Saravani[1], Lancelot Normand[1], Diego Maupomé[1], Fanny Rancourt[1], Thomas Soulas[1], Sara Besharati[1], Anaelle Normand[2], Sébastien Mosser[3] and Marie - Jean Meurs[1]

[1] *Université du Québec à Montréal, QC, Canada*

[2] *Université du Québec à Trois-Rivières, QC, Canada*

[3] *McMaster University, ON, Canada*

Abstract

Measuring the severity of the signs of eating disorders is a new task introduced in the eRisk 2022 competition. The goal of this task is to infer the severity of the signs of Eating Disorders (ED) in social media users given their writings on said platform. These signs are inventoried by a standard self-report questionnaire. We presented two unsupervised similarity-based models using per-question feature sets that were developed using expert knowledge of discussions of ED online. These features are supported by dense word representations, namely GloVe and word2vec. The obtained results were modest as no training data were available.

1. Introduction

Eating Disorders (ED) are characterized by the status of becoming so preoccupied with food and weight issues that a person can hardly focus on other aspects of their life [1]. ED, which might start with an obsession with food, body weight, or body shape [1], affect at least 9% of the population worldwide [2], and can cause serious, potentially fatal medical complications [1]. In recent years, there has been increasing interest in studying automatic detection of ED [3, 4, 5, 6, 7, 8, 9, 10] since patients with ED share similar linguistics and psychological features [2]. For example, CLEF eRisk[1] is a competition where the evaluation methodology, effectiveness metrics and practical applications of early risk detection in different mental health areas, including ED, are explored. Task 3 of eRisk 2022 [11] seeks to promote the development of automated means measuring the severity of the signs of ED in social media users as per the Eating Disorder Examination Questionnaire (EDE-Q) [12] based on their writings. This task follows in the spirit of eRisk 2018 Task 2, which was aimed at the early detection of signs of Anorexia [13].

CLEF 2022 – Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

hosseini_saravani.seyed_habib@courrier.uqam.ca (S. Hosseini Saravani); normand.lancelot@courrier.uqam.ca (L. Normand); maupome.diego@courrier.uqam.ca (D. Maupomé); rancourt.fanny.2@courrier.uqam.ca (F. Rancourt); soulas.thomas_david@courrier.uqam.ca (T. Soulas); besharati.sara@courrier.uqam.ca (S. Besharati); anaelle.normand@uqtr.ca (A. Normand); mossers@mcmaster.ca (S. Mosser); meurs.marie-jean@uqam.ca (M. -. J. Meurs)

[1] https://erisk.irlab.org/

This paper presents the similarity-based approaches proposed by the RELAI team to Task 3 of eRisk 2022. The proposed approaches rely on feature sets dedicated to each item in the questionnaire. The features sets developed with expert knowledge are compared to the written production of users based on pre-trained word vectors. The developed models try to measure the severity of the signs of ED in an unsupervised manner. The paper is structured as follows. Section 2 gives an overview of the relevant literature. Section 3 introduces the dataset and task in detail, while Section 4 presents the proposed approach. Finally, Sections 5 and 6 report on the experiments and conclude the paper.

2. Related work

Machine learning has been used for automatic detection of signs of eating disorders, using different data and approaches [4, 5, 8, 9, 10]. Astorino et al. [4] used a SVM classifier to analyze the relationship between nutrition, health status and well-being of individuals with the principal objective of identifying possible psychological elements associated with ED. They studied the situation of young people at an Italian high school regarding ED by carrying out a statistical survey on the students in relation to dietary habits, attitudes towards food and physical activity. Finally, the reported results suggest that machine learning algorithms can detect signs of ED based on socio-demographic and psychological characteristics. In a study by Haynos et al. [5], females with heterogeneous ED diagnoses completed demographic and psychiatric assessments at baseline, Year 1 and 2 follow-ups. Elastic net and logistic regression models were able to predict ED diagnosis, binge eating, compensatory behavior, and underweight Body Mass Index (BMI) at Years 1 and 2.

Separately, models based on similarity measures have been implemented for the detection of mental health disorders in social media. Maupomé et al. [8] presented similarity based models for early detection of signs of pathological gambling. In their approaches, they inferred whether a test user was at risk based on the similarity between the topics present in the textual productions of users were similar to those present in problem-gambling testimonials. In other work [9], comparing textual similarities was used to predict Beck Depression Inventory (BDI) answers for the users of social media. Hypothesizing that a person can produce contents similar to BDI answers, the cosine similarity between the encoded textual productions of the users and BDI answers was computed. In a similar work [10], cosine similarity was used to filter posts relevant to each BDI question. Through the use of pre-trained sentence transformer models, the proximity between user posts to each BDI answer was measured.

The present work sought then to apply the potential of similarity-based approaches to the assessment of ED signs from the writings posted on social media platforms.

3. Task and data

The eRisk 2022 Task 3 dataset has only a testing set containing 28 subjects. For each subject, there is a history of postings on the Reddit[2] social media platform and a questionnaire filled by

[2] https://www.reddit.com/r/EatingDisorders/

that subject. The questionnaire is extracted from the Eating Disorder Examination Questionnaire (EDE-Q)[3], which is a self-reported questionnaire with 28 items and adapted from the semi-structured interview Eating Disorder Examination (EDE) [12]. This questionnaire was designed to assess the range and severity of features associated with a diagnosis of eating disorder using four subscales (Restraint, Eating Concern, Shape Concern and Weight Concern) and a global score. However, only questions 1-12 and 19-28 were considered in this competition. The statistics of the eRisk 2022 Task 3 dataset are shown in Table 1.

Table 1
Statistics of the eRisk 2022 Task 3 data. Posts are considered per users.

Number of users	28
Minimum time span of posts in months	1
Maximum time span of posts in months	96
Average time span of posts in months	26.46
Minimum number of posts	15
Maximum number of posts	1548
Average number of posts	401.57

3.1. Evaluation metrics

In eRisk 2022 Task 3, evaluation is based on eight metrics: Mean Zero-One Error (MZOE), Mean Absolute Error (MAE), Macroaveraged Mean Absolute Error ($\mathrm{MAE}_{\mathrm{macro}}$), Restraint Subscale (RS), Eating Concern Subscale (ECS), Shape Concern Subscale (SCS), Weight Concern Subscale (WCS), Global ED (GED).

MZOE corresponds to the proportion of incorrect predictions from the system. MAE and $\mathrm{MAE}_{\mathrm{macro}}$ measure the average (resp. macro-averaged) deviation of the predicted response from the true response. Therefore, a perfect predictor would get scores of 0 for each of the metrics above. In the equations below, f denotes the classification done by an automatic system, Q is the set of questions of each questionnaire, q_i is the i-th question, $R(q_i)$ is the user's answer for the i-th question and $f(q_i)$ is the predicted answer of the system for the i-th question. Q_j also represents the set of questions whose true answer is $j \in \{0, 1, \dots, 6\}$.

$$\mathrm{MZOE}(f, Q) = \frac{|\{q_i \in Q : R(q_i) \neq f(q_i)\}|}{|Q|}$$

$$\mathrm{MAE}(f, Q) = \frac{\sum_{q_i \in Q} |R(q_i) - f(q_i)|}{|Q|}$$

$$\mathrm{MAE}_{\mathrm{macro}}(f, Q) = \frac{1}{7} \sum_{j=0}^{6} \frac{\sum_{q_i \in Q_j} |R(q_i) - f(q_i)|}{|Q_j|}$$

For MZOE, MAE, and $\mathrm{MAE}_{\mathrm{macro}}$, a single score is computed for each user, and the reported score is the average over all values.

[3] https://www.corc.uk.net/outcome-experience-measures/eating-disorder-examination-questionnaire-ede-q/

RS, ESC, SCS, and WCS are fine-grained metrics that refer to the four main concerns of people suffering from ED. For the restraint subscale, the associated ED score refers to the mean response to questions 1 through 5. Then, RS computes the Root Mean Square Error (RMSE) between the restraint ED score obtained from the user and the one obtained by the system. Similarly, ECS, SCS, and WCS are obtained from other subsets of questions—ECS is calculated for the questions (7, 9, 19, 20, 21), SCS is for questions (6, 8, 10, 11, 23, 26, 27, 28), and WCS is related to questions (8, 12, 22, 24, 25). For a set of users, U, and a model, f, the metric $S \in \{\mathrm{RS}, \mathrm{ECS}, \mathrm{SCS}, \mathrm{WCS}\}$ is computed as:

$$S(f, U) = \sqrt{\frac{\sum_{u_i \in U} (R_S(u_i) - f_S(u_i))^2}{|U|}}$$

where $R_S(u_i)$ is the ground-truth score for user u_i on the relevant subscale and $f_S(u_i)$ is the predicted score. Finally, GED corresponds to the RMSE between the real and an estimated global ED scores:

$$\mathrm{GED}(f, U) = \sqrt{\frac{\sum_{u_i \in U} (R_{\mathrm{GED}}(u_i) - f_{\mathrm{GED}}(u_i))^2}{|U|}},$$

where $R_{\mathrm{GED}}(u_i)$ and $f_{\mathrm{GED}}(u_i)$ are the total ground-truth and predicted scores for user u_i.

4. Methodology

Two similarity based models (each with four variations) and 22 feature sets designed using expert knowledge were developed. The remainder of this Section details how the questionnaires were automatically filled, as well as the feature sets and word representations on which this process relied.

4.1. Word representations

Two kinds of pre-trained word vectors were used as word representation. The first are 300-dimensional word2vec [14] word vectors trained on publicly available textual content such as Wikipedia and UMBC WebBase corpus [15] using the Continuous Bag Of Words (CBOW) model [16]. The second were GloVe [17] word vectors trained on two billion Twitter[4] posts (tweets). Unlike word2vec, GloVe does not rely only on local context words, but incorporates global word co-occurrence [14]. Three different sizes of GloVe word vectors were used in our models: 50, 100 and 200.

4.2. Feature sets

A set of features per question in the EDE-Q was developed using expert knowledge, totaling 22 feature sets. An expert identified words that would allow to answer a given question in the EDE-Q questionnaire. For example, for the question *In the last 28 days, have you felt fat?* (Q11), words such as *fat* and *scale* were considered as features. As a result, a set of around 30 features

[4] https://twitter.com/

per question was prepared (for details, see Table 6 in Appendix A). After developing the feature sets, each feature was associated with the relevant pre-trained word vector, resulting in a set of word vectors per question.

4.3. Models

In accordance with the use of the EDE-Q (see Section 3), the models examined only the writings posted in the last 28 days by a given user. These writings are combined into a single text document. Then, for a given question in the questionnaire, only the set of words in the history present in the associated feature set were kept; all other words were removed. As with the feature sets, these words were replaced by the relevant word vectors, resulting in 22 sets of vectors per user, each pertaining to a different question in the questionnaire.

Model-A In Model-A (named RELAI_A in the eRisk 2022 report), the word vectors in each feature set were aggregated together by averaging, yielding one vector per question. The same was done to the sets of word vectors selected for each users, obtaining 28 vectors each representing a user. Next, for each user, the cosine similarity between the vector representing that user and the ones developed for 22 questions was computed, obtaining 22 similarity values per user.

$$\mathrm{CosSim}(q_i, u_i) = \frac{q_i^\top u_i}{\|q_i\| \|u_i\|},$$

where q_i, u_i refer to the feature (resp. user) vectors corresponding to the i-th question of the questionnaire. As no training data were available to empirically select thresholds, intervals associated to each answer were roughly of the same length ($\approx 1/7$). The scores are defined are as follows:

$$\mathrm{score}(q_i, u_i) = \begin{cases} 0, & \text{if } \mathrm{CosSim}(q_i, u_i) \leq 0.142; \\ 1, & \text{if } 0.142 < \mathrm{CosSim}(q_i, u_i) \leq 0.284; \\ 2, & \text{if } 0.284 < \mathrm{CosSim}(q_i, u_i) \leq 0.426; \\ 3, & \text{if } 0.426 < \mathrm{CosSim}(q_i, u_i) \leq 0.568; \\ 4, & \text{if } 0.568 < \mathrm{CosSim}(q_i, u_i) \leq 0.710; \\ 5, & \text{if } 0.710 < \mathrm{CosSim}(q_i, u_i) \leq 0.852; \\ 6, & \text{otherwise.} \end{cases} \tag{1}$$

Model-B In Model-B (named RELAI_B in the eRisk 2022 report), like in Model-A, the word vectors in each feature set were aggregated together by averaging, yielding one vector per question, but the same was not done to the sets of word vectors selected for each users. Instead, given a user, the similarity between the vectors obtained per question and each of the pre-trained word vectors selected for the user was computed separately, obtaining a list of similarity values. By averaging the similarity values, a single value was obtained representing the similarity of the user with the question. Finally, a score per question was attributed as prescribed in (1) to predict the final score , *i.e.* $\mathrm{score}(q_i, u_i)$.

4.4. Evaluation on proxy data

Since eRisk 2022 Task 3 dataset only has testing data, we used the eRisk 2018 dataset [13] for early detection of signs of anorexia to evaluate the performance of our models. This dataset contains a collection of Reddit posts from people who may or may not suffer from an eating disorder. The original dataset contains a set of Reddit posts per subject (sometimes spanning multiple years). For the purposes of the EDE-Q, a dedicated dataset was created from the aforementioned dataset. Each subject's history of writings was then sliced into periods of 28 days and treated as a separate observation. This sought to reduce the bias for annotators: having only a period of 28 days to go on, we ensured the the EDE-Q questionnaire was filled with only data from the period of 28 days imposed by the questionnaire.

Filling the questionnaires. The process described above resulted in a set of 362 observations. They were distributed to two annotators with different levels of expertise according to the presence of the curated features mentioned previously. Each post was analyzed individually, computing the number of features present. Posts were marked as potentially containing ED-related discourse if: (1) if different feature words comprised at least 25% of the post length (in words) or, (2) over 10 feature words were present. The latter condition was added to contend with short posts.

Observations where at least one post fulfills at least one of these conditions were assigned to a psychology student with deep knowledge on online discussions about ED. The remaining documents – the ones deemed unlikely to contain any mention of ED – were assigned to an annotator with no expert knowledge on ED. The annotators were instructed to solely rely on the information contained in the document to estimate the answer that the given subject would have recorded on the questionnaire. For example, in a document where a subject detailed their diet and calorie intake over the course of the month, the annotator could make an informed assessment for the aspect of ED addressed in question 1: *Have they been deliberately trying to limit the amount of food they eat to influence their shape or weight (whether or not they have succeeded)?*

Models and evaluation results on proxy data. Based on the type and size of the word vectors used, we obtained eight models as shown in Table 2. In the first experiments on proxy data, the feature sets with all the features were used in the models. Obtained results are shown in Table 3. However, assuming that more precise feature sets with smaller intersection with each other can improve the results, the feature sets sizes were reduced to 10 words on average. This process was done in two steps: (1) removal of adverbs of frequency (e.g. always, often, sometimes), (2) keeping the words that were closely associated to a given question. As shown in Table 4, reducing the number of features improved the results.

Three observations can be made from the results obtained on the proxy dataset: (1) word vectors with higher dimensions improved the results, (2) for each model configuration, Model-B slightly outperformed Model-A, (3) the models with more specific features had better results. Accordingly, the following models were selected for the competition: Model-A-W2V, Model-B-W2V, Model-B-W2V-AllFeatures, Model-B-GloVe200, Model-B-GloVe200-AllFeatures.

Table 2
Models developed

Name	Approach	Type of embedding	Dimension
Model-A-W2V	Model-A	word2vec	300
Model-A-GloVe50	Model-A	GloVe	50
Model-A-GloVe100	Model-A	GloVe	100
Model-A-GloVe200	Model-A	GloVe	200
Model-B-W2V	Model-B	word2vec	300
Model-B-GloVe50	Model-B	GloVe	50
Model-B-GloVe100	Model-B	GloVe	100
Model-B-GloVe200	Model-B	GloVe	200

Table 3
Results on proxy data using all the features (lower is better)

Model	MZOE	MAE	GED	RS	ECS	SCS	WCS
Model-A-W2V	0.24	0.44	1.00	1.21	0.86	1.00	0.95
Model-A-GloVe50	0.29	0.57	1.19	1.44	1.03	1.10	1.19
Model-A-GloVe100	0.27	0.52	1.11	1.36	0.95	1.05	1.09
Model-A-GloVe200	0.26	0.49	1.07	1.30	0.92	1.02	1.05
Model-B-W2V	**0.23**	**0.41**	**0.98**	**1.16**	**0.83**	**0.98**	**0.93**
Model-B-GloVe50	0.28	0.54	1.15	1.37	0.99	1.07	1.15
Model-B-GloVe100	0.27	0.49	1.08	1.29	0.93	1.03	1.05
Model-B-GloVe200	0.26	0.47	1.04	1.24	0.90	1.00	1.01

Table 4
Results on proxy data using more specific features (lower is better)

Model	MZOE	MAE	GED	RS	ECS	SCS	WCS
Model-A-W2V	**0.11**	0.25	0.87	1.04	**0.69**	0.92	0.83
Model-A-GloVe50	0.12	0.28	0.93	1.07	0.72	1.01	0.92
Model-A-GloVe100	0.12	0.27	0.90	1.05	0.72	0.96	0.87
Model-A-GloVe200	0.12	0.26	0.89	1.04	0.70	0.95	0.86
Model-B-W2V	**0.11**	**0.24**	**0.86**	**1.03**	**0.69**	**0.91**	**0.82**
Model-B-GloVe50	0.12	0.28	0.92	1.06	0.72	0.99	0.90
Model-B-GloVe100	0.12	0.26	0.89	1.04	0.70	0.96	0.86
Model-B-GloVe200	**0.11**	0.26	0.88	**1.03**	0.70	0.94	0.85

5. Experimental results

As the results presented in Table 5 show, the best performances were achieved by Model-B-W2V-AllFeatures and Model-B-GloVe200-AllFeatures. Hence, the performances recorded at the test phase are significantly worse than those on the proxy data. Furthermore, unlike the results obtained on the proxy dataset, the best performing models at the eRisk competition used the less specific feature sets. The limited generalization capabilities of the proposed models suggest that some level of overfitting occurred. In addition, among the four subscales, the recorded results on the Eating Concern Subscale (ECS) are consistently better than that of the other subscales, which suggests that our features better captured some discourse elements of eating concerns.

Table 5
Results on the eRisk 2022 Task 3 obtained by the selected models (lower is better)

Model	MZOE	MAE	$\mathrm{MAE}_{\mathrm{macro}}$	GED	RS	ECS	SCS	WCS
Model-A-W2V	**0.82**	3.31	2.91	3.59	3.65	3.05	4.19	3.74
Model-B-W2V	**0.82**	3.32	2.91	3.59	3.66	3.05	4.19	3.74
Model-B-W2V-AllFeatures	**0.82**	3.19	2.74	3.34	3.15	2.80	4.08	3.64
Model-B-GloVe200	**0.82**	3.30	2.89	3.56	3.65	3.03	4.17	3.71
Model-B-GloVe200-AllFeatures	0.83	**3.15**	**2.70**	**3.26**	**3.04**	**2.72**	**4.04**	**3.61**

6. Conclusion and future work

For measuring the severity of the signs of Eating Disorders, we presented five models that used two main algorithms based on similarity measurement. In addition, we developed a feature set for each of the 22 questions (topics) of the questionnaire. The results achieved were moderate since it was a new challenge without training data. Future work will aim to use word vectors trained on data more germane to the setting to represent user posts and the feature sets. This can be used in combination with unsupervised machine learning algorithms.

Reproducibility. The source code of the proposed approaches is licensed under the GNU GPLv3.

Acknowledgments

This research was enabled in part by support provided by Calcul Québec and the Digital Research Alliance of Canada. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [MJ Meurs, NSERC Grant number 06487-2017], and the support of the Government of Canada's New Frontiers in Research Fund (NFRF), [MJ Meurs, NFRFE-2018-00484].

References

[1] NAMI: Eating Disorders, https://www.nami.org/About-Mental-Illness/Mental-Health-Conditions/Eating-Disorders/Overview, 2022. (accessed: 2022.05.15).

[2] V. Cuteri, G. Minori, G. Gagliardi, F. Tamburini, E. Malaspina, P. Gualandi, F. Rossi, M. Moscano, V. Francia, A. Parmeggiani, Linguistic feature of anorexia nervosa: A prospective case–control pilot study, Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity 27 (2022) 1367–1375.

[3] S. B. Wang, Machine learning to advance the prediction, prevention and treatment of eating disorders, 2021.

[4] A. Astorino, R. Berti, A. Astorino, V. Bitonti, M. D. Marco, V. Feraco, A. Palumbo, F. Porti, I. Zannino, Early detection of eating disorders through machine learning techniques, in: International Conference on Learning and Intelligent Optimization, Springer, 2020, pp. 33–39.

[5] A. F. Haynos, S. B. Wang, S. Lipson, C. B. Peterson, J. E. Mitchell, K. A. Halmi, W. S. Agras, S. J. Crow, Machine learning enhances prediction of illness course: a longitudinal study in eating disorders, Psychological medicine 51 (2021) 1392–1402.

[6] J. Fardouly, R. D. Crosby, S. Sukunesan, Potential benefits and limitations of machine learning in the field of eating disorders: current research and future directions, Journal of Eating Disorders 10 (2022) 1–14.

[7] S. Sadeh-Sharvit, E. E. Fitzsimmons-Craft, C. B. Taylor, E. Yom-Tov, Predicting Eating Disorders from Internet Activity, International Journal of Eating Disorders 53 (2020) 1526–1533.

[8] D. Maupomé, M. D. Armstrong, F. Rancourt, T. Soulas, M.-J. Meurs, Early detection of signs of pathological gambling, self-harm and depression through topic extraction and neural networks, Proceedings of the Working Notes of CLEF (2021).

[9] D. Maupomé, M. D. Armstrong, F. Rancourt, M.-J. Meurs, Leveraging Textual Similarity to Predict Beck Depression Inventory Answers, in: Proceedings of the Canadian Conference on Artificial Intelligence, 2021.

[10] D. Inkpen, R. Skaik, P. Gamaarachchige, D. Angelov, M. T. Fredenburgh, UOttawa at eRisk 2021: Automatic filling of the Beck Depression Inventory questionnaire using deep learning, Working Notes of CLEF (2021) 21–24.

[11] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, erisk 2022: Pathological gambling, depression, and eating disorder challenges, in: European Conference on Information Retrieval, Springer, 2022, pp. 436–442.

[12] C. Fairburn, Z. Cooper, M. O'Connor, Eating Disorder Examination (Edition 16.0 D), Cognitive behavior therapy and eating disorders (2008) 265–308.

[13] D. E. Losada, F. Crestani, J. Parapar, Overview of eRisk: Early Risk Prediction on the Internet, in: International conference of the cross-language evaluation forum for european languages, Springer, 2018, pp. 343–361.

[14] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, A. Joulin, Advances in Pre-Training Distributed Word Representations, in: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), 2018.

[15] L. Han, A. L. Kashyap, T. Finin, J. Mayfield, J. Weese, UMBC_EBIQUITY-CORE: Semantic Textual Similarity Systems, in: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity, 2013, pp. 44–52.

[16] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[17] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

A. Feature sets developed for the questions of the eRisk questionnaire

Table 6: Appendix A: Feature sets developed for the questions of the eRisk questionnaire

Q	Features
1	Suppressed, Restrict, Less, Too much, Skinny, Fat, Minimum, Maximum, Fail, Hungry, Fed up, Satiated, Empty, Water, Hurt, Tired, Crave, Shame, Portions, Tiny dishes, Carbs, Calorie, Surplus, Diet, Fasting, Sometimes, Often, Always, Never, Scales, Add, Remove, Light, Fill
2	Fasting, Empty, Endure, Craving, Stomach Rumbling, Skip, Meals, Water, Irritable, Faint, Sleep, Tired, Avoid, Waiting, Time, Distract, Sport, Busy, Changing mind, Weigh, Self-control, Goal, Challenge, Intermittent fasting, Hurt, Skip, Motivation
3	Meat, Sugar, Carbs, Restraining, Privation, Starving, Craving, Strict, Diet, Rules, Have to, Pleasure, Fat, Desire, Exception, Comfort food, Taste, Goal, Avoid, Flavor, Envy, Self-Control, Jealous, Prevent, Groceries, Compare, Nutritional values, Insipid, Light
4	Rules, Limitation, Strict, Calories, Count, Compare, Goal, Weight, List, Cheat, Bad, Stress, Pressure, Challenge, Severely, Measures, Control, Watching, Numbers, Choose, Diet, Low fat, Sugar free, Low carbs, No dressing, Light, Healthy food, Advice, Impose, Small portions
5	Empty, Full, Heavy, Free, Lightweight, Active, Restrictions, Water, Substitute, Control, Busy, Awake, Strict, Rumbling, Stomach, Burning, Dynamic, Skinny, Tired, Confidence
6	Curves, Compare, Skinny, Model, Social medias, Beach, Summer, Sport, Abs, Restrict, Beauty, Confidence, Standards, Society, Measure, Weight, Internet, Magazines, Effort, Waist, Flat, Hungry, Progress, Mirror, Shape, Tiny, Small, Feminine
7	Numbers, Distract, Obsession, Priority, Wasting time, Focus, Others, Daydreaming, Frustration, Relate, Link, Avoid, Occasions, Calculate, Run away, Busy, Attention, Concerned, Envy, Craving, Break, Brain
8	Complex, Compare, Ugly, Fat, Others, Focus, Hide, Avoid, Uncomfortable, Attention, Disgust, Concerned, Yes, No, Sometimes, Obsession, Anxiety, Daydreaming, Shame, Shy
9	Control, Anxiety, Limitation, Body, Thinking, Stress, Goals, Strategies, Ration out, Small portion, Advice, Motivation, Scared, Shame, Self, Concentrated, Sick, Restrictions, Too much
10	Fat, Ugly, Weight, Measure, Privation, Diet, Fat pad, Size, Anxiety, Compare, Numbers, Skinny, Ambition, Progress, Insecure, Scales, Obesity, Body, Starving, Overweight, Afraid, Oversize clothes, Sugar, Sport, Control, Snacks, Size, Bullying, Vomit, Complex
11	Insecure, Yes, Sometimes, Often, Ugly, Shame, Obese, Bulge, Sometimes, Eating, Meals, Imperfections, Body, Skinny, Less, More, Scales, Weight, Stress, More, Less, Afraid, Diet, Compare, Goal, Fat
12	Sport, Calories, Skinny, Progress, Fat, Envy, Obsession, Diet, Count, Weight, Compare, Progress, Portions, Water, Hungry, Proud, Tired, Motivation, Success, Failure, Lbs, Kilos, Skip meals, Starving, Concentration, Goal, Numbers, Scales, Discouraged, Rules
19	Cheat, Same, A lot, Pleasure, Comfort food, Watching, Embarrassed, Hiding, Fat, Sugar, Lot, Good, Excitation, Guilt, Taste, Bit, People, Family, Bad, Flavor, Bedroom, Scared, Heavy, Lie, Allowed, Appetite

20	Guilt, Often, Bad, Too much, Quantities, Cook, Sugar, Carbs, Strict, Cheat, Sport, Water, Diet, Throw up, Sad, Rules, Hurt, Weak, Allowed, Angry, Anxious
21	Shy, Embarrassed, Ugly, Judgement, Mocking, Concerned, Hiding, Cook, Meals, Healthy, Pretend, Compare, Hungry, Disgust, Slow, Bites, Tiny dishes, Portions, Control, Oppressed, Home, Dirty, Messy
22	Healthy, Perception, Self-esteem, Judgement, Criticism, Prejudices, Confidence, Mental, Body, Expectations, Insecure
23	Self-esteem, Compare, Judgement, Mocking, Beliefs, Able to, Confidence, Limits, Affect, Perception, Expectations, Body, Insecure
24	Stress, Waiting, Goal, Progress, Deception, Scared, Upset, Realize, Afraid, Ashamed, Weight, Lbs, Kilos, Numbers, Heavy, Light
25	Heavy, A lot, Diet, Fat, Fat pad, Weight, Scales, Sizes, Clothes, Large, Mocking, Shame, Envy, Jealousy, Slack, Compare, Frustration, Ugly, Pressure, Insecure, Skinny, Obese, Disgust, Mirror, Flat , Thinner, Healthy
26	A lot, Better, Curves, Skinny, Fat, Model, Sport, Standard, Social medial, Disgust, Shame, Ugly, Embarrassed, Compare, Gym, Flat, Slack, Frustration, Envy, Suffer, Tummy
27	Myself, Sadness, Shame, Disgust, Uncomfortable, Imperfections, Details, Body, Embarrassed, Slack, Fat, Reflection, Bad, Hiding, Watching, Scared, Suffer, Unsure, Underweight, Overweight
28	Insecure, Comparison, Watching, Uncomfortable, Shame, Shy, Hiding, Avoid, Stress, Jealousy, Envy, Observed, Clothes, A lot, Tight, Social anxiety, Bad, Sad, Embarrassed, Toilet, Large, Disgust, Mocking, Scared, Body, People, Pretend