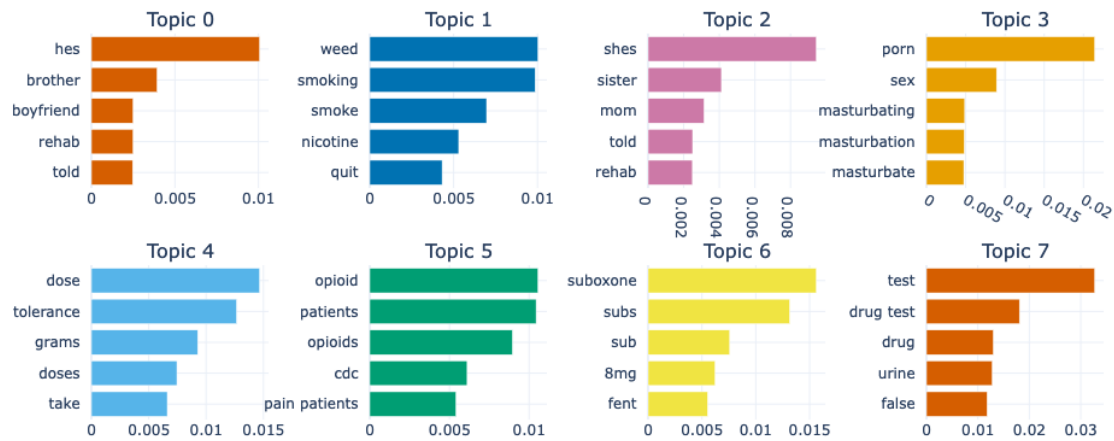


# Dynamic Topic Modeling of Kratom Use and Experiences: Insights on 13 years of Reddit Discussions

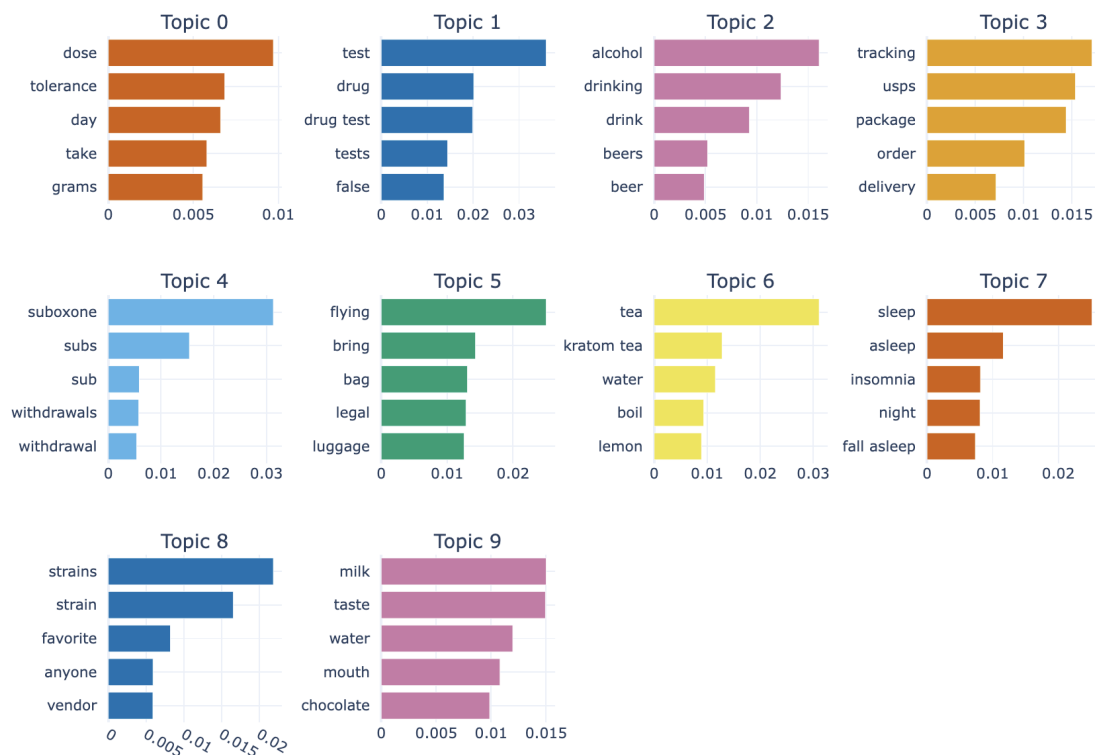
## Supplementary Materials

| Module (purpose)                                                                          | Hyperparameter                     | Value                                                     |
|-------------------------------------------------------------------------------------------|------------------------------------|-----------------------------------------------------------|
| Embedding model<br>(convert Reddit posts into numerical representations i.e., embeddings) | English Sentence transformer model | all-MiniLM-L6-v2                                          |
| UMAP model<br>(dimensionality reduction of extracted embeddings)                          | Number of neighbours               | 15                                                        |
|                                                                                           | Number of components               | 5                                                         |
|                                                                                           | Minimum distance                   | 0.1                                                       |
|                                                                                           | Random state                       | 42                                                        |
| HDBSCAN model (cluster embeddings of similar posts)                                       | Minimum cluster size               | 15                                                        |
| CountVectorizer (tokenize topics)                                                         | Stop words                         | NLTK english stop words + ['http', 'https', 'amp', 'com'] |
|                                                                                           | Ngram range                        | (1,2)                                                     |
| Class TF-IDF model<br>(create topic representation of each cluster of similar posts)      |                                    | Default parameters                                        |

**Table S1.** Hyperparameters set for each module of BERTopic, including a brief description of the purpose of each module. Most default values are retained from instantiating BERTopic out-of-the-box.



**Figure S1.** Results from a preliminary analysis of the dataset that included posts not filtered to only contain mentions of ‘kratom’ from the subreddits *r/addiction*, *r/chronicpain*, and *r/opiatesrecovery*. The figure depicts the Top 5 representative words based on class-based term frequency-inverse document (c-TF-IDF) scores for the top 8 topics that emerged. The higher the c-TF-IDF score, the more relevant the word is in the context of a given topic. The results show that noise is introduced to the extracted topics when not filtering the posts specifically for kratom (e.g., topics emerging that are related to sex addiction, marijuana).



**Figure S2.** Results from the sensitivity analysis of posts from kratom-specific subreddits only (i.e., excluding posts from *r/addiction*, *r/chronicpain*, *r/opiatesrecovery*). The ChatGPT-4o generated topic labels are as follows: 0 -

Tolerance Break Strategies for Managing Daily Dosing, 1 - Concerns about Kratom and False Positives in Drug Tests, 2 - Using Kratom to Manage Alcohol Cravings, 3 - USPS Package Tracking Issues, 4 - Using Kratom to Manage Suboxone Withdrawal, 5 - Traveling with Kratom: Tips and Experiences, 6 - Kratom Tea Recipes and Preparation Techniques, 7 - Effects of Kratom on Sleep Patterns, 8 - Kratom Strain Varieties and Vendor Practices, 9 - Methods for Consuming Kratom with Milk and Water.