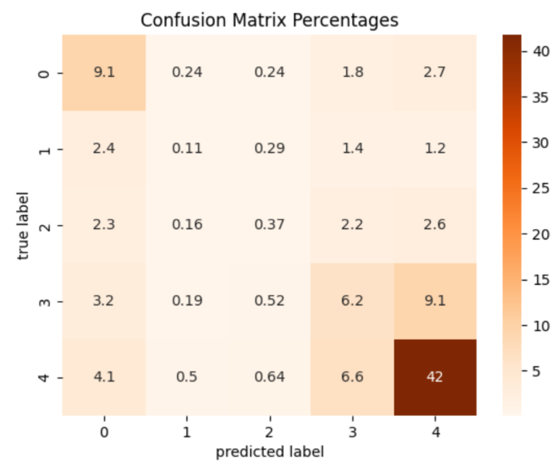
2.14.23 - Katin Lind - katinansa@proton.me

Psychedelic Efficacy Problem Statement

How well might psychedelic drugs work to treat mental illness, compared with prescription psych meds?
What insights can be drawn from anonymously-submitted psychedelic experience reports?

Summary of Findings



A ComplementNB model was trained on 31,559 patient-provided reviews and ratings of prescription psych meds from a medical study. The model was then used to predict ratings that would be assigned to 4,562 narrative psychedelic experience reports submitted (without quantitative ratings included) to erowid.org. On average, the predicted efficacy rating of psychedelic drugs was, on a scale of 1-5, 4.49, compared with a 3.93 actual average rating of prescription drugs.

The ComplementNB model was selected as the best learner across multiple iterations of a randomized grid search, performing better on training data than other classifiers and better than a regression model which would have returned ratings on a scale of 0.0-4.0 rather than categories of 0, 1, 2, 3, and 4 as the classifier did. A naive model resulted in an F1 score of 0.53 and log loss of 16.76, while the final model's evaluation metrics were as follows: 0.58 F1 score, 1.27 log loss, 0.75 roc_auc, and accuracy with best k=2 of 0.73.

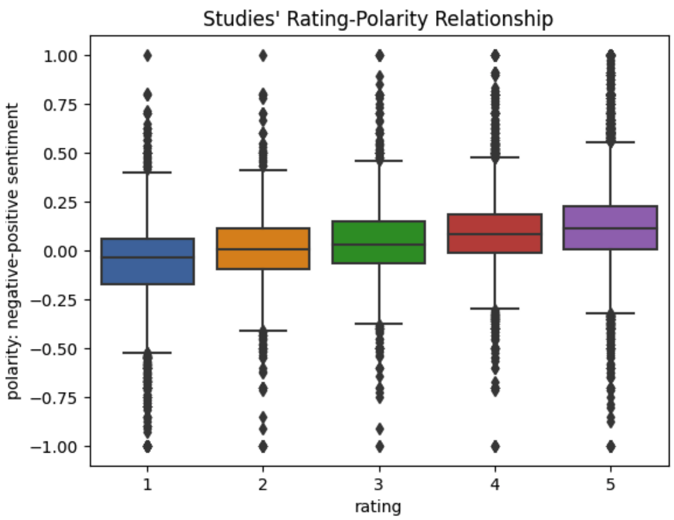
The confusion matrix demonstrates the percentage of labels that fall into each true:predicted category; this dataset was not grossly unbalanced but definitely skewed toward more ratings being perfect 5/5 (4/4 in the context of the ML algorithm), so most labels were correctly predicted to fall in the higher end of the range, and missed predictions were more likely to be assigned a score that was too high. Another way of interpreting the results could be as follows:

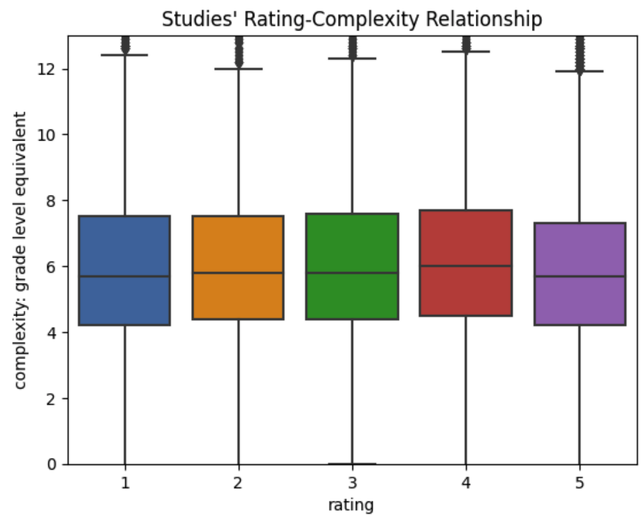
When the model predicted a score of…	4	3	2	1	0
It was correct this percent of the time…	73	34	18	9	43
And within 1 this percent of the time…	89	82	57	43	55

Process

A note on the overall methodology: typically, text features from a language-based product review would not be used to predict a product rating, as a single number is likely easier to gather than a full text. However, due to criminalization of psychedelic drugs, there are relatively few formal scientific studies of psychedelic efficacy. There are a large number of narrative reports wherein people share their experiences, but these typically do not include quantitative ratings and are often submitted anonymously to internet forums. The unconventional methodology of predicting ratings based on the contents of a review has enabled me to bring this wealth of data about psychedelics into a format where criminalized drugs can more easily be compared quantitatively with prescription drugs.

To train the model, first the texts of prescription drug reviews were cleaned with the following goals in mind: Punctuation was largely removed, but emoji and symbols clearly associated with sentiment i.e. '!' were kept. The standard English stopword list from spacy was revised to remove words that likely indicate a drug user's opinion i.e. 'not.' Remaining stopwords were removed, and texts were run through spelling correction and lemmatization.

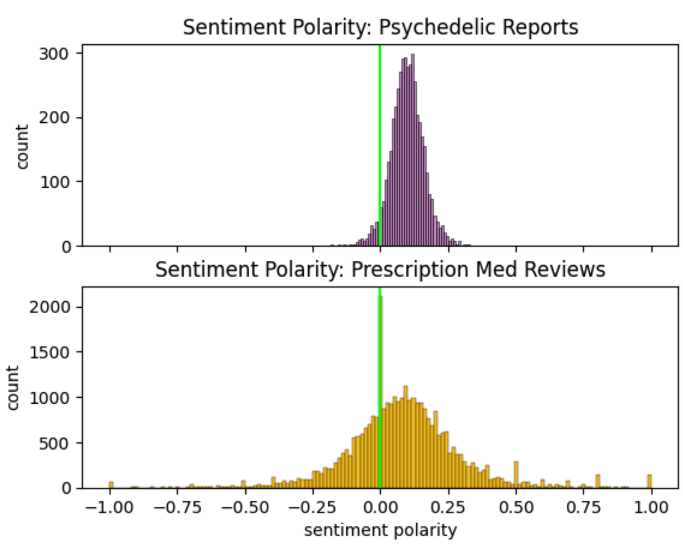




Several features were engineered. Not all were strongly correlated with the target feature "rating," but during model tuning, options that included feature reduction performed less well than models that maintained all of the following features: text character length; text complexity using flesch-kincaid grade levels; subjectivity and polarity based on TextBlob sentiment classification; and a feature I named "similarity_w_10," which indicates how similar each individual review's language is to a meta-review of all perfectly-rated drugs, according to spacy's text similarity function. The figures above demonstrate the degree of correlation between the dependent variable and the most- and second-most correlated features, polarity and complexity.

Once features were engineered, a pipeline was constructed including a ColumnTransformer along with the ComplementNB classifier. Features listed above were normalized with min-max scaler, and the actual review texts were transformed using a CountVectorizer, resulting in a sparse matrix.

After predictions were made and the model was evaluated as described in the "Summary of Findings" section, the trained model was applied to unseen data, which consisted of psychedelic experience reports scraped from Erowid and pre-processed in the same manner as the original reviews. Predicted ratings were adjusted based on inaccuracies observed during original model evaluation, detailed in the table above indicating "When the model predicted a score of. . ." For example, 73% of ratings predicted "4" were maintained as such, a random 16% of "4"s were switched to "3," etc. From here, comparisons could be made with the original prescription drug ratings.



Again, the results were as such: the average psychedelic drug rating was 4.49, and the average prescription drug rating was 3.93. This makes sense, given that the feature most correlated with ratings in the training data, "polarity," does appear to vary in value across the datasets, with psychedelic reports generally having less negative sentiment.

Concluding Recommendations

The results of this report indicate that psychedelic drugs may be favored by many users over prescription drug options such as benzodiazepines or SSRIs for feeling emotionally well and treating such common conditions as depression, addiction, or anxiety.

The findings should, however, be considered in context. The prescription drug ratings were gathered via a more formal study, and although Erowid does strive to represent a range of experiences including "bad trips," reports drawn from this source may come from people biased toward psychedelic enthusiasm. Furthermore, the model, while significantly outperforming a naive model, could be improved upon considerably to be more sensitive especially to the text features of reviews for low- and moderately-rated drugs. A more robust model may more reliably predict meaningful ratings for the unseen data of psychedelic reports.

Although this report provides just one, imperfect data point in comparing psychedelic and prescription drugs, it does at the very least indicate promise in pursuing further research. I would encourage all sorts of healthcare organizations, from policy advocacy groups, to drug development startups, to clinical research facilities, to explore options for providing and continuously evaluating psychedelic treatment options.

Additional Considerations

As mentioned, I would like to improve model performance–and, potentially, generalizability to unseen data–by switching from CountVectorizer to word embeddings as a method for quantifying raw text data. This could enable the model to more accurately make new predictions even when words do not exactly match those of the original training data's reviews.

Furthermore, there are opportunities to expand the application of this model. So far, it has only been applied to psychedelic experience reports scraped from erowid.com. Other forums exist, such as Psychonaut Wiki or Reddit. It is impossible to know before checking, but it may be that the format and distribution of key features such as polarity in reports from these sources may more closely match the features of the training data, which could result in more confidence in predicted ratings. I would like to move on to scrape, clean, and apply the model with narratives from these additional sources.

Beyond assigning ratings, many opportunities exist now that all of these narratives have been processed. Key words unique to reviews associated with high ratings could be analyzed to better understand drug features that people appreciate in either prescription or psychedelic medicines. So far, I have identified common words and n-grams present in each dataset, but I would like to further differentiate the following representation by drug rating.

More common in psychedelic reports	In both sets' top 20 lemmas	More common in prescription reviews
come, experience, friend, get, hour, try, know, look, much, no, thing, think, trip,	effect, feel, go, like, start, take, time	! , anxiety, day, depression, help, life, meditation, mg, month, sleep, week, work, year