

Psychedelic Efficacy

Problem Statement:

How might we predict and improve the efficacy of psychiatric treatment with a prescription psych med or psychedelic option, given demographic information about a patient and/or narrative description of their experiences before and after trying a drug?

Context:

Most trials of prescription psych meds are rightfully conducted in controlled experiments where patients carefully report details about their experiences so that researchers can evaluate the effects of a drug for specific treatment purposes. Psychedelics are a popular form of medicine for mental health and represent a growing edge of psychiatric research, but studies are still relatively limited in scope compared with those for more conventional prescription drugs.

There do exist, however, extensive narrative reports shared across Erowid, Psychonaut Wiki, and Reddit, informally evaluating the effects of psychedelics. A model trained on data from a structured study can be applied to summarized unstructured data to predict the treatment efficacy of psychedelic medicines. This can open extensive, under-utilized data dating back decades to constructive use in a medical context.

Criteria For Success

- Predict the quantitative efficacy of a prescription psych med based on only a patient's narrative description of their experience with the drug. This comes from structured data that includes demographic information, narrative reports of experiences, and a numeric, ordinal score evaluating the drug's efficacy in helping somebody improve their emotional well-being.
- Summarize unstructured narratives scraped from psychedelic experience reports. Extract demographic information and information about the submitter's history with mental illness or emotional wellness goals where possible.
- Assess unstructured narrative data for viability to be processed using the model trained on labeled data. Where possible, predict quantitative labels of psychedelics' efficacy.
- Clearly evaluate and explain likely limitations that exist when comparing narratives from studies versus anonymous forum submissions.

Scope and Constraints:

Predicting quantitative scores of a prescription drug based on accompanying descriptions of subjective experience will require careful text processing but is a fairly straightforward machine learning task given well-labeled data that can be split into train and test sets.

Significant differences may exist between data gathered via a study versus that scraped from a website. Psychedelic experience reports are often quite long and unstructured. Because psychedelics are only recently being decriminalized in some places in the U.S, reports are often submitted anonymously and include limited background information about the user. Submissions are often made to forums intended for those who are curious about trying drugs for help with mental health challenges or simply for fun. The narratives are often written from peer-to-peer, rather than from patient-to-researcher as done in formal studies. My hypothesis is

that psychedelic reports may tend toward relaying more extremely positive or extremely negative experiences, since users voluntarily submit narratives when they feel they have something meaningful to share. This would be in contrast with narratives shared during a controlled study where a fuller range of experiences may be expected.

These characteristics do not make psychedelic experience reports a less meaningful source of information about the effects of a drug—subjective experiences of emotional changes matter when it comes to evaluating medicine for mental health—but these differences should be kept in mind when interpreting predictions across data sources.

A note on the term “psychedelic”: It does not have a single agreed-upon definition. When people use terms such as “plant-based medicine,” “hallucinogen,” or “entheogenic substance” (drug that increases feelings of social or spiritual connectedness), they may be referring to psychedelics. For the purposes of this research, psychedelics will be defined as substances that have some degree of consciousness-altering, hallucinogenic effect and that are more often considered entheogenic than deliriant (deliriant are hallucinogens that are almost always described as having nightmarish effects). Specifically, this research will focus on three chemical classes: tryptamines (LSD, psilocybe mushrooms, DMT, and related substances), phenethylamines (psychedelic amphetamines including mescaline, MDMA, or 2-CX), and arylcyclohexylamines (psychedelic dissociatives including PCP, PCE, or Ketamine)

Stakeholders

- MAPS (Multidisciplinary Association for Psychedelic Studies) advocates for decriminalization and medical use of psychedelics and may be able to use my report as one source of information in their advocacy efforts.
- As psychedelics gain cultural acceptability with decriminalization, many healthcare startups will likely wish to invest in psychedelic research, development, and distribution. This has been happening in the booming cannabis industry already. As psychedelic popularity grows, businesses will need to make wise choices about which specific substances to promote for which purposes, and the results of this research will be invaluable.
- I will be scraping psychedelic experience reports from websites that millions of people use when trying to find information about psychedelics that they wish to try. Each narrative as-is can be helpful in preparing people for what to expect when consuming a substance, but narratives are not easily searchable. I will be processing, summarizing, and labeling the narratives in ways that could make them more user-friendly for people everywhere.

Data Sources

The following datasets consist of tabular data with dependent variables that include overall rating for the efficacy of a prescription psych med in treating mental illness as well as other labels such as efficacy for specific symptoms like racing thoughts or depressed mood. Independent variables include columns with a drug's name, patient demographic information, and narrative descriptions of drug effects:

- <https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Druglib.com%29>
- <https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29>
- <https://www.askapatient.com/store/#!/Psytar-Data-Set/p/449080512/category=129206256>

Psychedelic experience reports will be scraped from the following sources:

- <https://erowid.org/experiences/>
- https://psychonautwiki.org/wiki/Experience_index

- Numerous sub-reddits, selected from among those listed here:
<https://www.reddit.com/search/?q=psychedelics&type=sr>

Preliminary Plan

This project will consist of two major stages.

First, tabular data will need to be cleaned. This will include wrangling data from multiple studies whose results were shared in a similar format including demographic information, narratives, and quantitative labels. This will be followed by natural language processing which will involve, at least, any feature engineering necessary to conduct sentiment analysis.

Degree of positive or negative sentiment expressed in descriptions of drug effects will become one of the independent variables in creating a predictive model, along with any other language features that appear to be meaningful within the narratives as well as demographic information that already exists within the tabular data.

A model will be trained that can predict the numeric rating of a drug's efficacy in treating symptoms of mental illness. The model could stand alone to predict efficacy of certain drugs for certain patients with any given demographic characteristics. Modeling should also be conducted, however, using only the processed narrative features as independent variables, because unstructured data gathered later may not include demographic data.

Insights about how to process narrative data related to psychiatric drug trials could be drawn from these similar studies::

- <https://zenodo.org/record/55013#.XPE1MC1eN24>
- <https://zenodo.org/record/3236318>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7447679/>

The second stage will begin with scraping unstructured psychedelic reports into a format where natural language processing and sentiment analysis can be conducted. Narratives' language features should be compared with those from the structured studies, and feature engineering should be conducted in such a way as to result in independent variables that match those expected by the predictive model. It may be that many scraped narratives will need to be discarded or truncated in order to reasonably apply the model and make predictions about quantitative labels of psychedelics' effects.

Work done for this project can be viewed here: https://github.com/fractaldatalearning/psychedelic_efficacy. I am particularly eager to point out that this project contains not just notebooks but packages of functions and unit tests in the 'src' folder.