Psychedelic Trip Report Generator

A Flatiron Phase 5 Project by Jordan Loewen-Colón

Today’s Agenda

BUSINESS PROBLEM

GOALS

DATA

METHODS

RESULTS

The Problem

The MindSpectrum Research Institute is working to **discover new medicines.**

The Major Challenge: is a scarcity of firsthand accounts or personal descriptions of these experiences.

Our Objective: leverage existing personal narratives and use artificial intelligence (AI) to **create realistic 'trip reports' categorized by different types of drugs**.

Primary Goals

Develop a text generator model

Aim Produce human-like trip reports

Focus on minimizing false positives

The Dataset

Before Cleaning

70k+ entries

Report length varies from 3 to 32k words.

Other data points include:

- Drug (11728 unique entries)
- Dosage
- Delivery
- Weight
- Year

After Cleaning and Engineering

67875 entries

10 Drug categories based on the Alcohol and Drug Foundation labels

Processed Report column:

- Lemmatized
- Tokenized
- Lower-cased

Methods

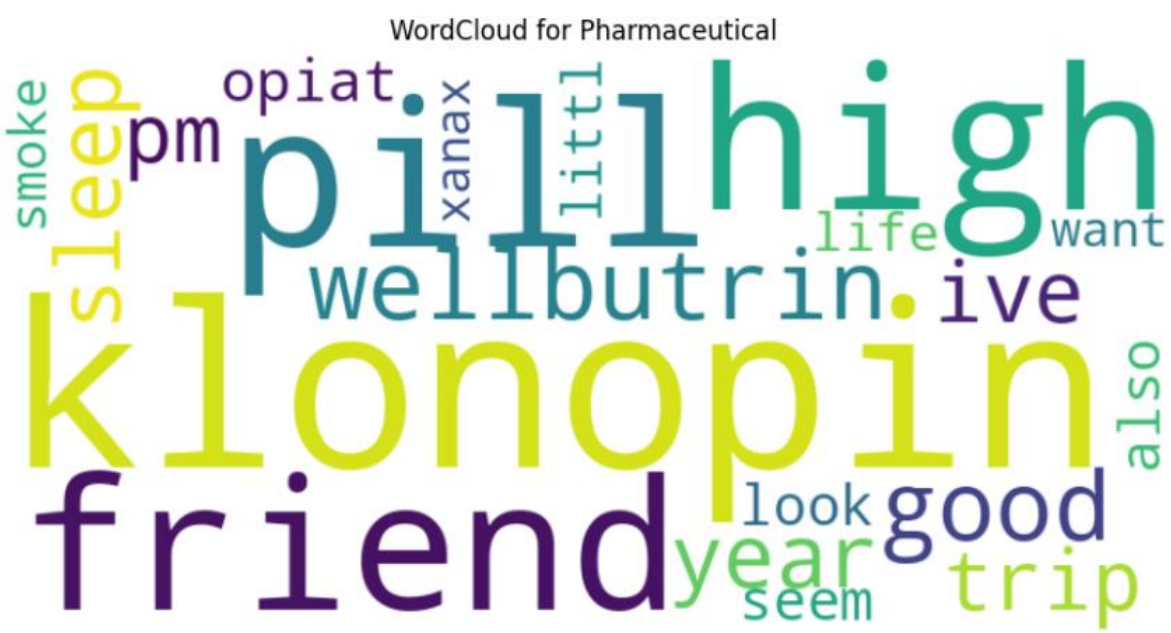
Vectorization and Unsupervised Learning To Discover Patterns

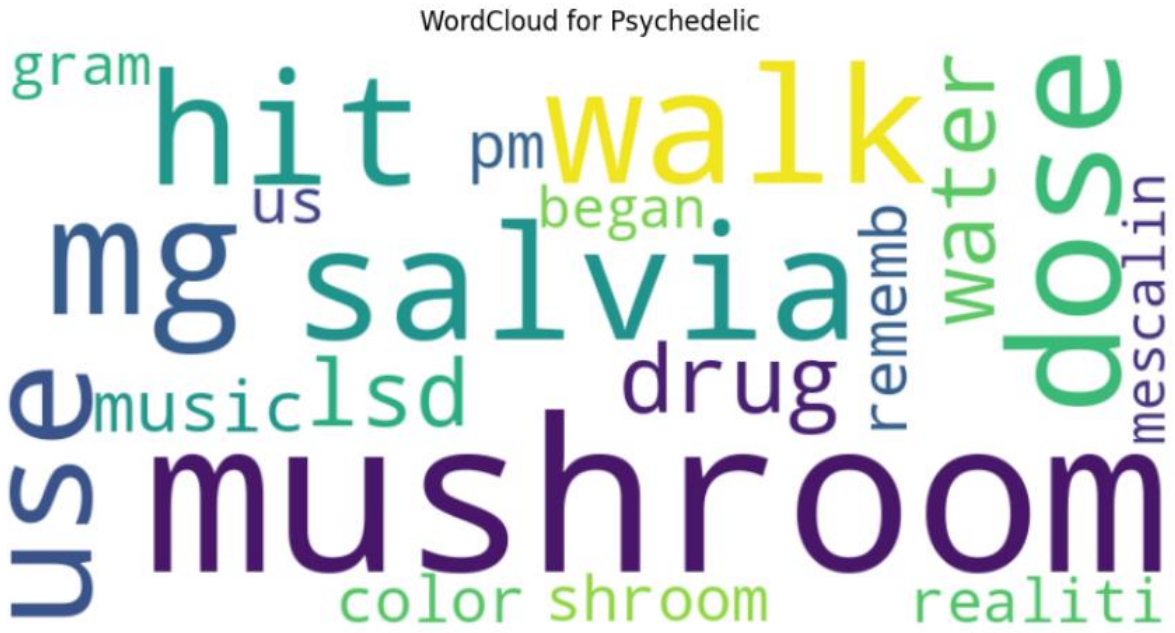
Multiclassification for drug prediction

Large Language Model to help verify text output

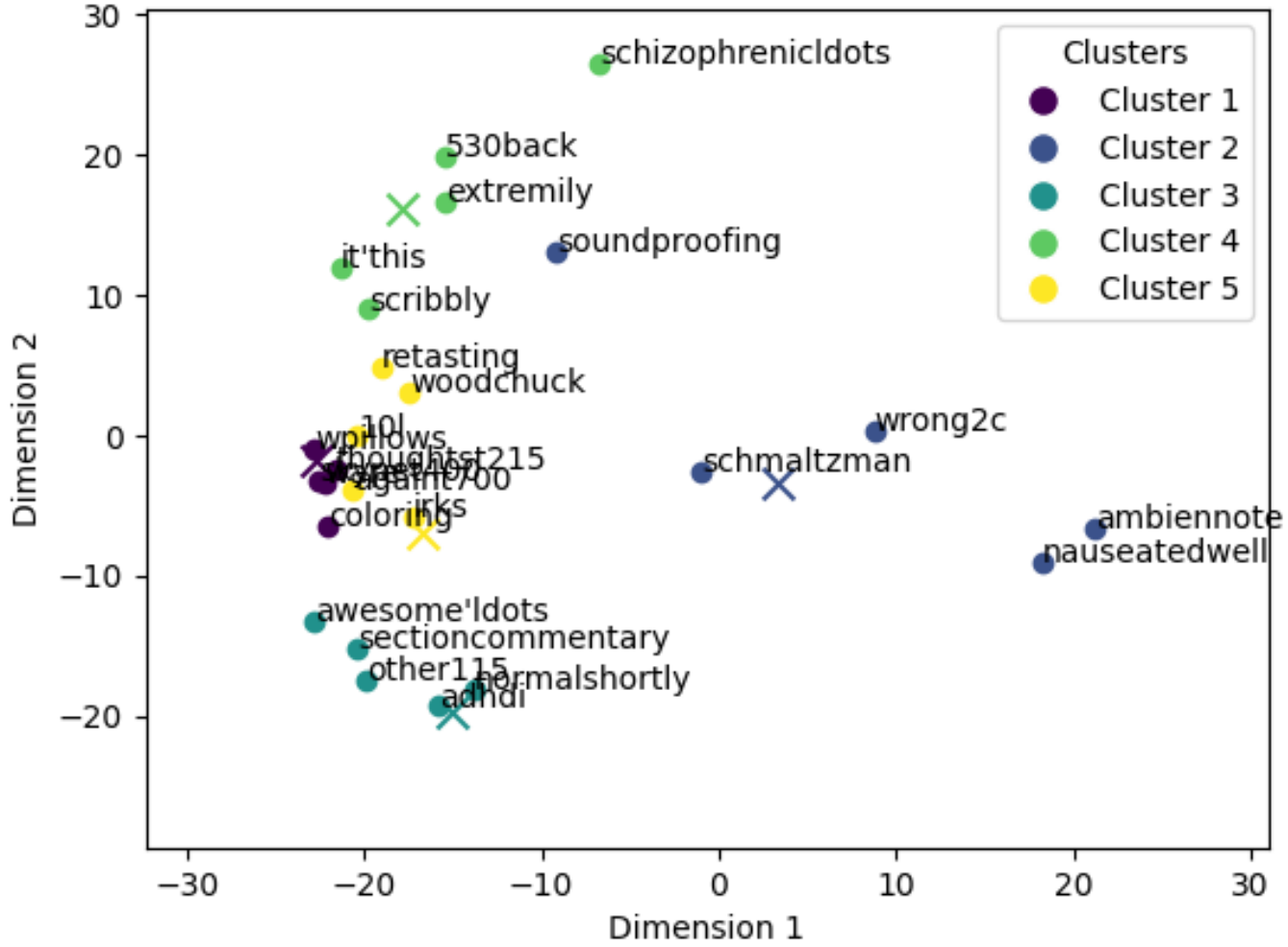
GPT-2 for Text Generation

Word Clouds from Unsupervised Learning Model (LDA)





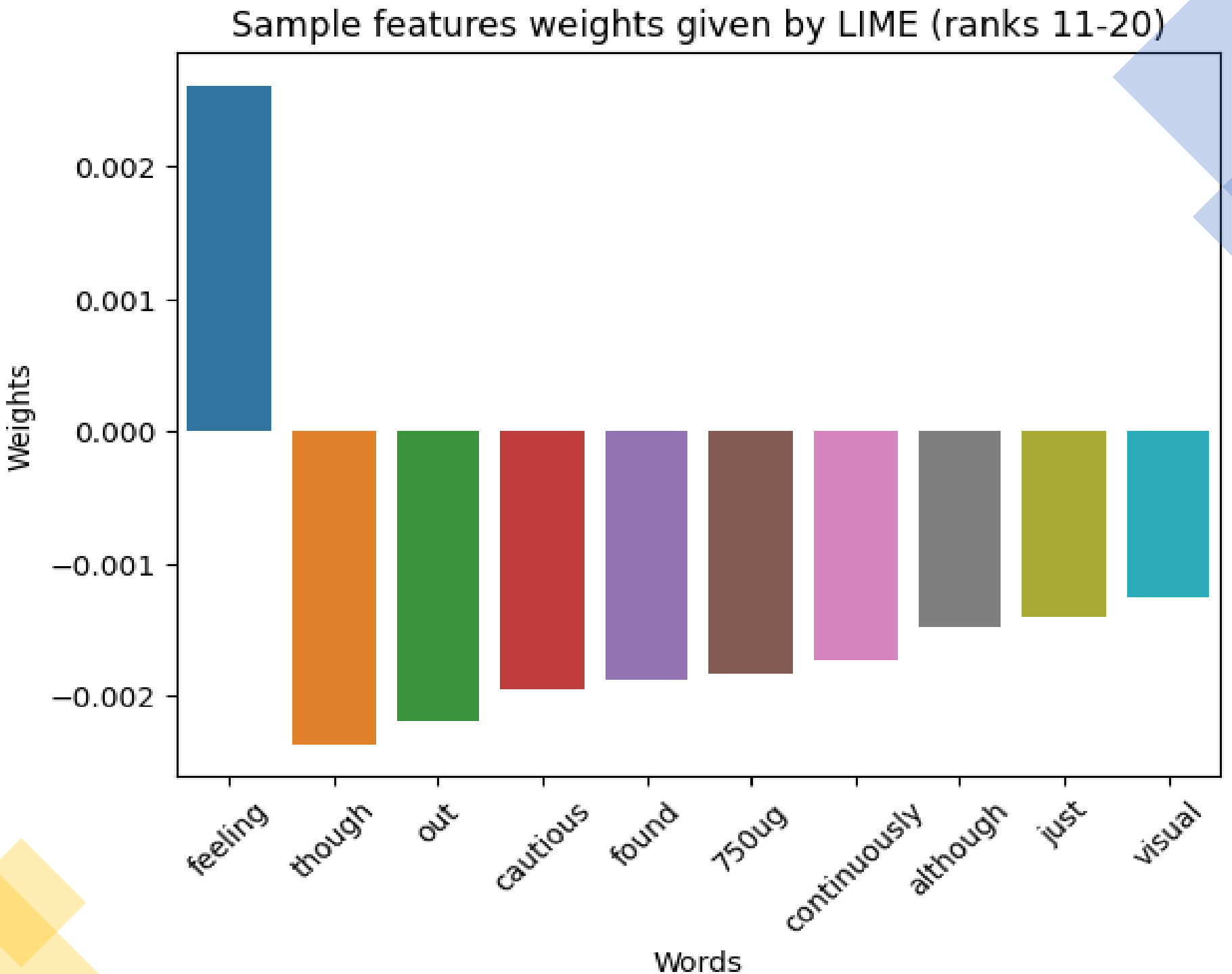
The model was able to highlight major differences in thematic content for each drug-sub-category



- 5 Clusters
- No clear connection, pattern, or theme
- Word combinations contributing “noise” to our data processing!

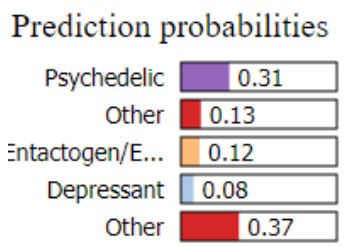
The weight of importance for how a word would contribute to a positive prediction of “Psychedelic”

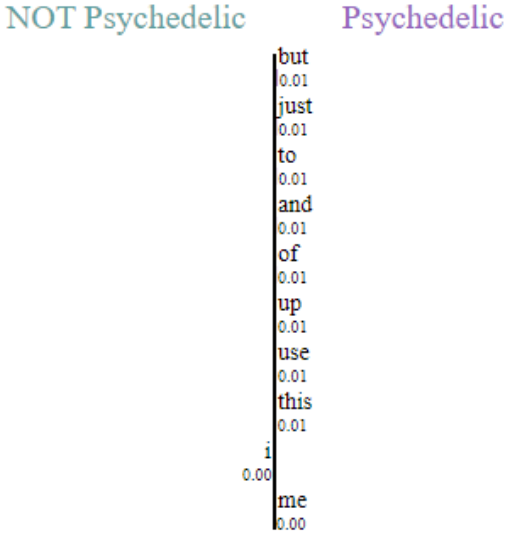
The words **“feeling”** and **“though”** are the most weighty, but inversely.



LIME Explainer Report

Unfortunately:
The model was too distracted with “stop words”

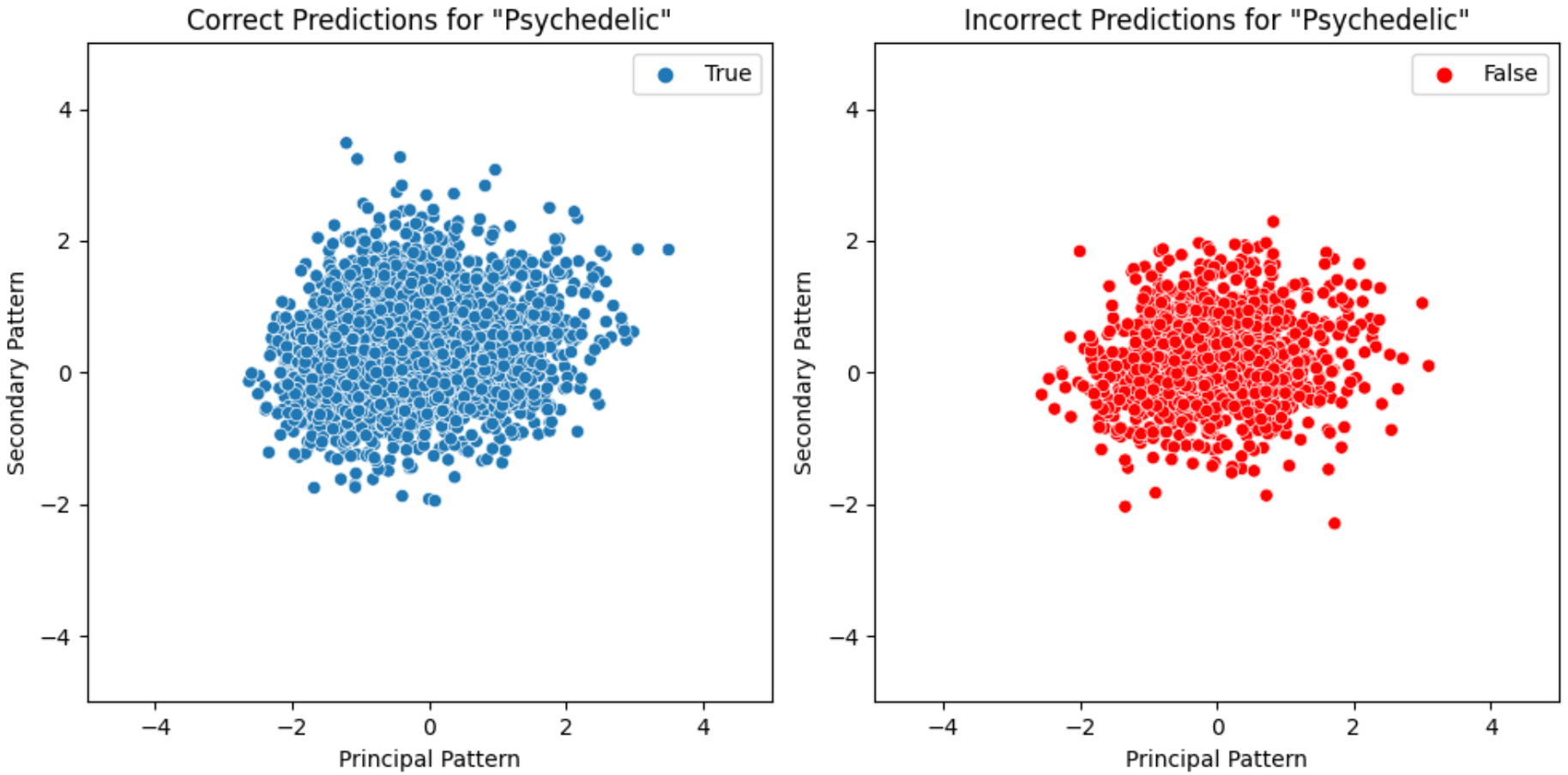




Text with highlighted words

first i would like to say that i am somewhat seasoned explorer " of the mind i have use many substance for purpose of recreation or exploration i 'm not boasting or anything i just want to be thorough i have been known to be able to keep my cool in le than desirable phsychedelic condition i have watched many friend cause long term damage to their psyche by misusing psychedelics and i try to remember this and repect the chemical i use i just want to imply that i can not compare the salvia experience to anything else i wa truly terrified like never before in my life the day of my last experience my girlfriend and i were scheduled to leave for week vacation the following day so there wa greatly relaxed atmosphere i wa in great mood maybe slightly apprehensive i had only smoked marijuana and am regular smoker i wa not stoned at the time i had experienced mild effect from salvia 10x but i did n't hold it in before but nothing like the one i wa about to i think this led to over eagerness and lack of respect i just did n't think it wa this

Underlying Patterns

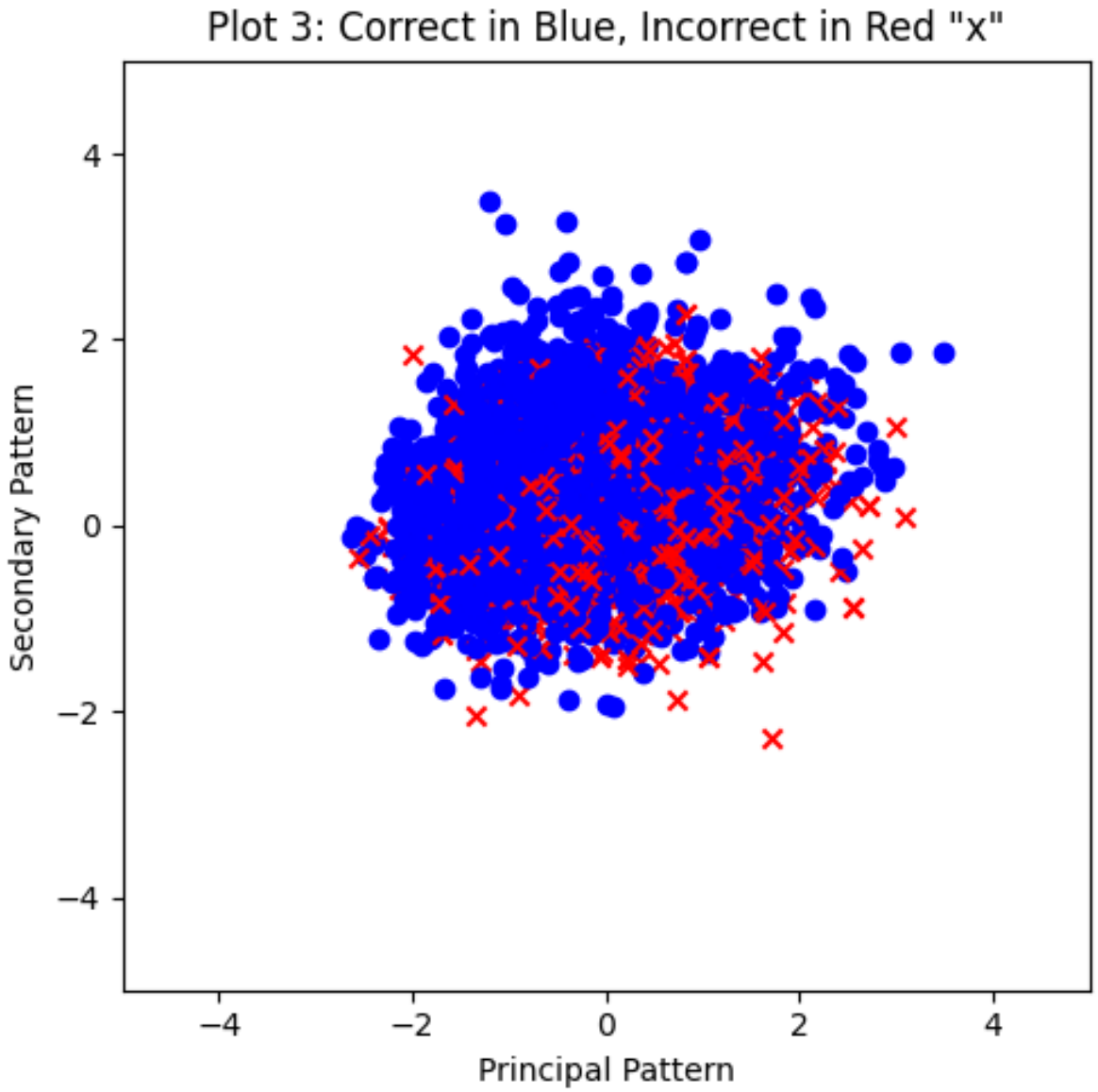


Positive predictions track with the “secondary pattern” and negative predictions track with the “principal pattern”

The difference is encouraging!

The model is picking up on something that allows it to actively differentiate.

More time and data might tell us what that difference is



Conclusions

Significance of the word "Feeling” (and stop words)

Improvement with Larger Data Sample for Psychedelics
(more data leads to higher precision)

Underlying Patterns in Predictions
(More time could allow for more understanding)

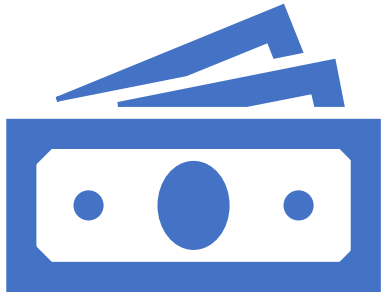
Recommendations

Experiment with removing stop words from reports to see if the model is able to focus more clearly on key words in reports.

Incorporate BigBird feature embeddings into GPT-2 model and use the trained RFC model to validate GPT-2 outputs.

Concerns & Interest

Computational efficiency limited pattern recognition



Findings indicate there is value in further study!

Thank You

Jordan Loewen-Colón

jbloewen@syr.edu