

OCD and Anxiety Diagnosis

By: Alec Edgecliffe-Johnson

Table of Contents

01

Purpose

What are we here to do today?

02

Data

Discussion of data sources and gathering

03

Methodology

How we arrived at our model

04

Models

Discussion of our models and model performance

05

Streamlit

Demo of Streamlit App

06

Further Steps

What we'd like to do in the future

Anxiety Disorders v. Obsessive Compulsive Disorders

OCD

"Obsessive-compulsive disorder (OCD) features a pattern of unwanted thoughts and fears (obsessions) that lead you to do repetitive behaviors (compulsions)." – Mayo Clinic

Anxiety

"...people with anxiety disorders frequently have intense, excessive and persistent worry and fear about everyday situations. Often, anxiety disorders involve repeated episodes of sudden feelings of intense anxiety and fear or terror that reach a peak within minutes (panic attacks)." – Mayo Clinic

Three Core Issues



Increased cases

According to the NIH $\frac{1}{3}$ people aged 13 - 18 are predicted to have an anxiety disorder in the coming years. Furthermore, COVID is increasing fears for many OCD patients.

Access to help

Mental health is a growing concern for most western health systems and it can be difficult to access quality medical care - particularly in certain parts of the world where there is less awareness.

Stigma

A stigma around mental health makes it difficult to speak about which can lead to lower rates of diagnosis and treatment.



Our Purpose

"Leverage anonymized, free-text posts made in an open, safe environment from individuals discussing obsessive compulsive or anxiety disorders in order to build a first-step (non-medical) diagnostic tool."

Strategy

1. Preparing Data

Gathered and cleaned data from pushshift reddit api.

3. Modeling

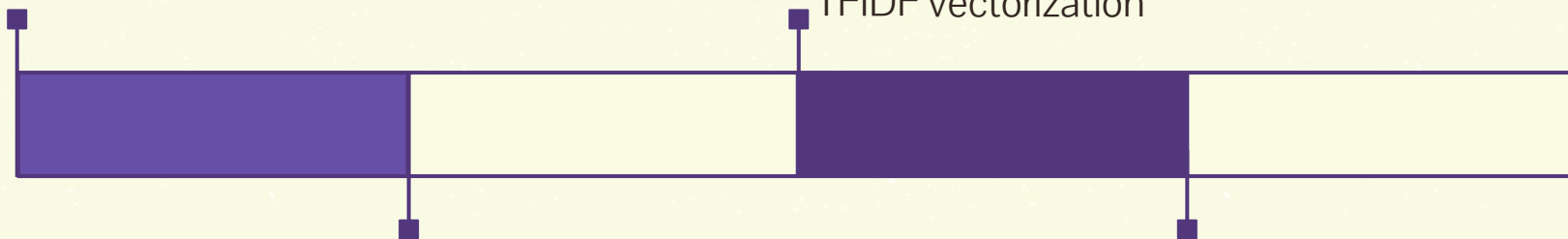
Created three gridsearch model pipelines with CV and TFIDF vectorization

2. EDA

Conducted exploratory data analysis to inform general findings

4. Streamlit

Build first-step non-medical diagnostic Streamlit application





01 Preparation

Source of, and method of
acquisition for, our data

Data



r/Anxiety
437K members

"Discussion and support for sufferers and loved ones of any anxiety disorder."

r/OCD
110K members



"A subreddit dedicated to discussion, articles, and images regarding OCD."

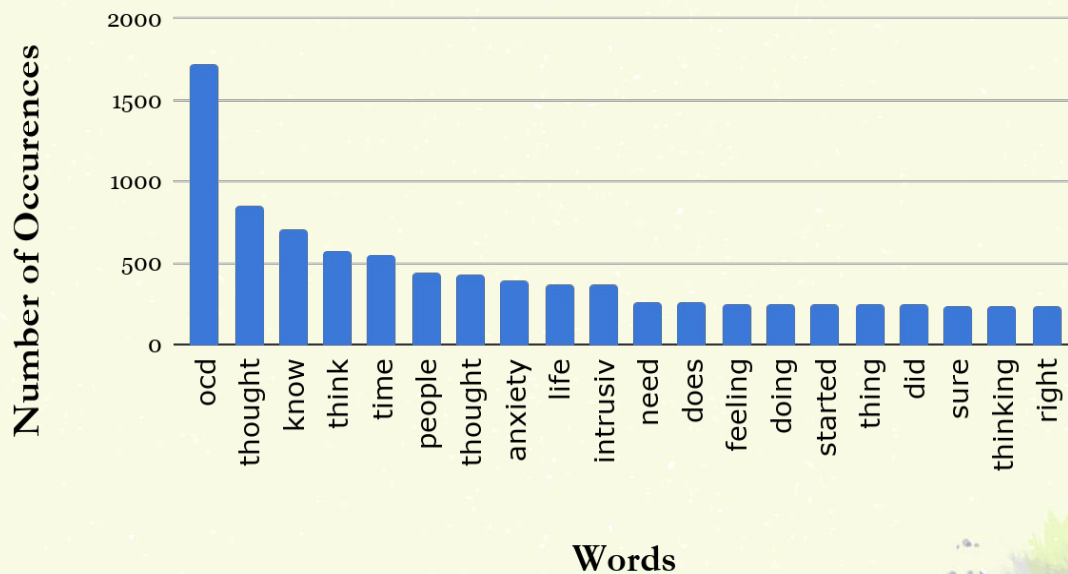


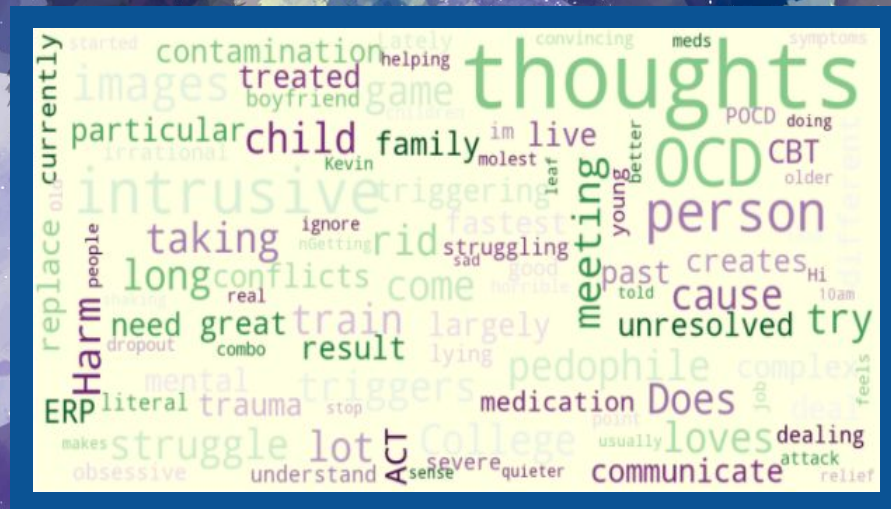
02 EDA

Word frequencies and
sentiment analysis insights

Word Frequencies – OCD

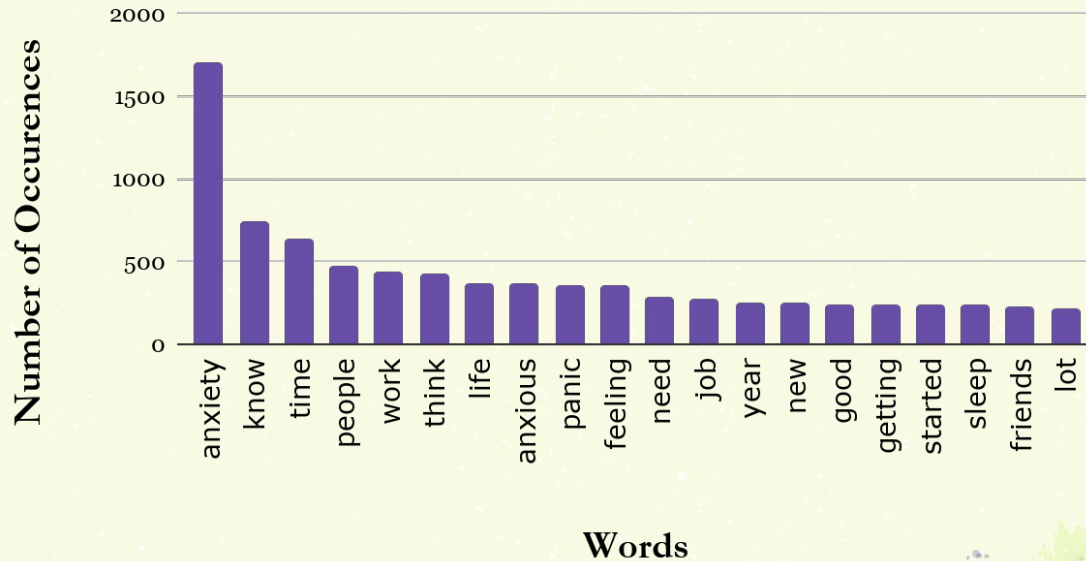
20 Most Frequent Words - OCD





Word Frequencies - Anxiety

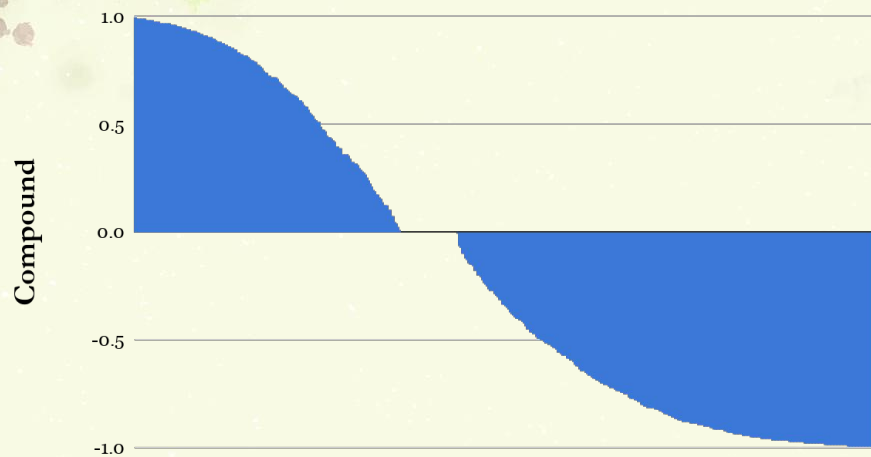
20 Most Frequent Words - Anxiety





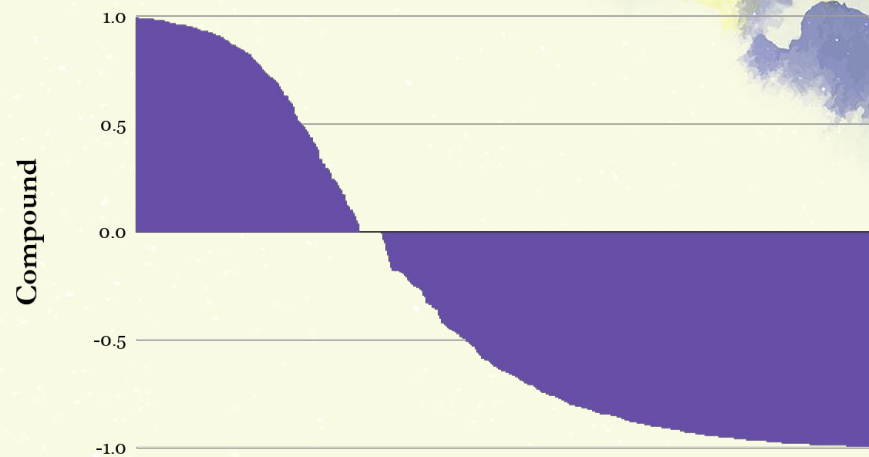
Sentiment Analysis – OCD v. Anxiety

Sentiment - OCD



Mean: -.18

Sentiment - Anxiety



Mean: -.30



03 Modeling

Develop three models and
assess misclassifications

Modeling

01

Baseline

- .50 predictive accuracy

02

Logistic Regression

- **Cvec** - .996/.913
- **Tvec** - .956/.908

03

Multinomial Bayes

- **Cvec** - .912/.88
- **Tvec** - .926/.870

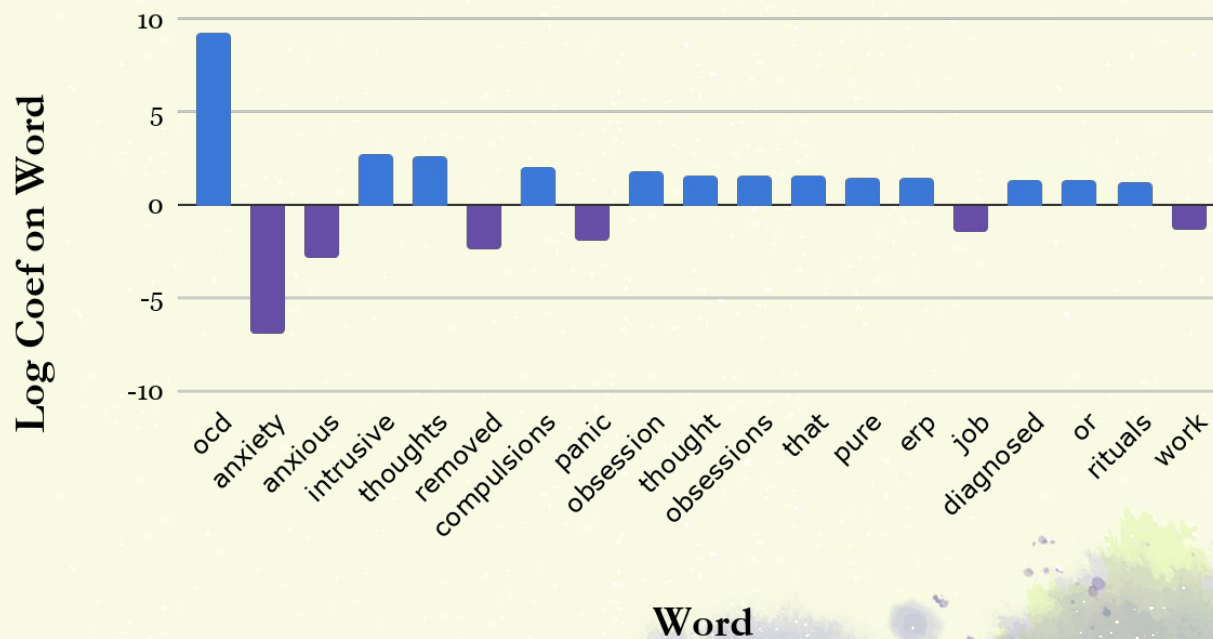
04

Random Forest

- **Cvec** - .998/.905
- **Tvec** - .998/.900

Log Reg Model Word Importance

20 Most Important Words



Log Reg (tvec) Confusion Matrix

	Actual Positive	Actual Negative
Predicted Positive	283	14
Predicted Negative	41	262

Log Reg Word Misclassification

20 Most Frequent Words in Misclassified Posts





04 Streamlit

First-step non-medical
diagnostic tool

Further Steps

1

Accuracy

Training on more data
and more steps taken
in preprocessing
(lemmatization)

2

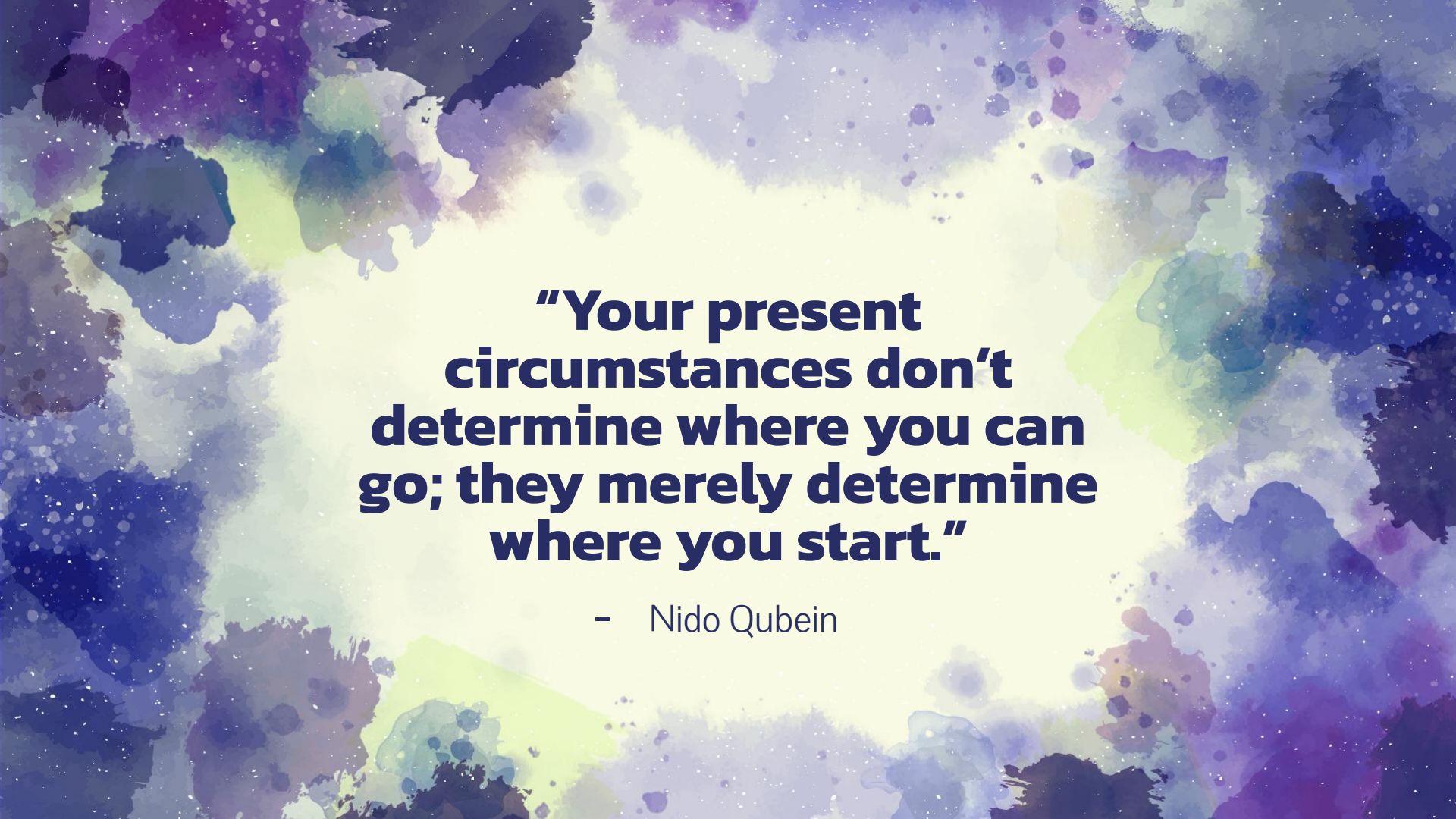
User and professional input

Input from doctors and
experts as well as
diagnosed individuals.
Perhaps reddit URLs?

3

Further Applications

Using mental health
focused reddit groups to
broaden use cases



**"Your present
circumstances don't
determine where you can
go; they merely determine
where you start."**

– Nido Qubein



Thank you!

Any questions?

Slide template by SlidesGo