



# Machine Learning Guild

## Capstone Presentation

► Emerging Tech and Tech Leadership

Presented by:  
Swapna Batta

# About me



**Swapna Batta**

## Professional Background

- Background in Engineering and Business Analytics
- Data Analytics Manager
- Completed 4 years at Deloitte

## Expertise

- Design and development of analytic tools/solutions in Audit Analytics
- Leading analytic projects in various industries - Healthcare, Process & Industrial Products(P&IP) and Financial Services Industry(FSI).

## Fun fact

- I enjoy learning Kathak (Indian dance form), playing table tennis and guitar when I get off work.
- I like reading comics and Medium blog posts

# AI in Psychotherapy

Psychotherapy is a general term for treating mental health problems by talking with a psychiatrist, psychologist or other mental health provider.

**1 in 5** U.S. adults experience mental illness each year  
**1 in 25** U.S. adults experience serious mental illness each year  
**1 in 6** U.S. youth aged 6-17 experience a mental health disorder each year  
**50%** of all lifetime mental illness begins by age 14, and 75% by age 24  
Suicide is the **2nd leading** cause of death among people aged 10-34



- **Drive for AI in Psychotherapy**
  - Psychotherapy is expensive
  - Comfort in expression
  - Availability in urgent moments
  - Accountability during therapy progression

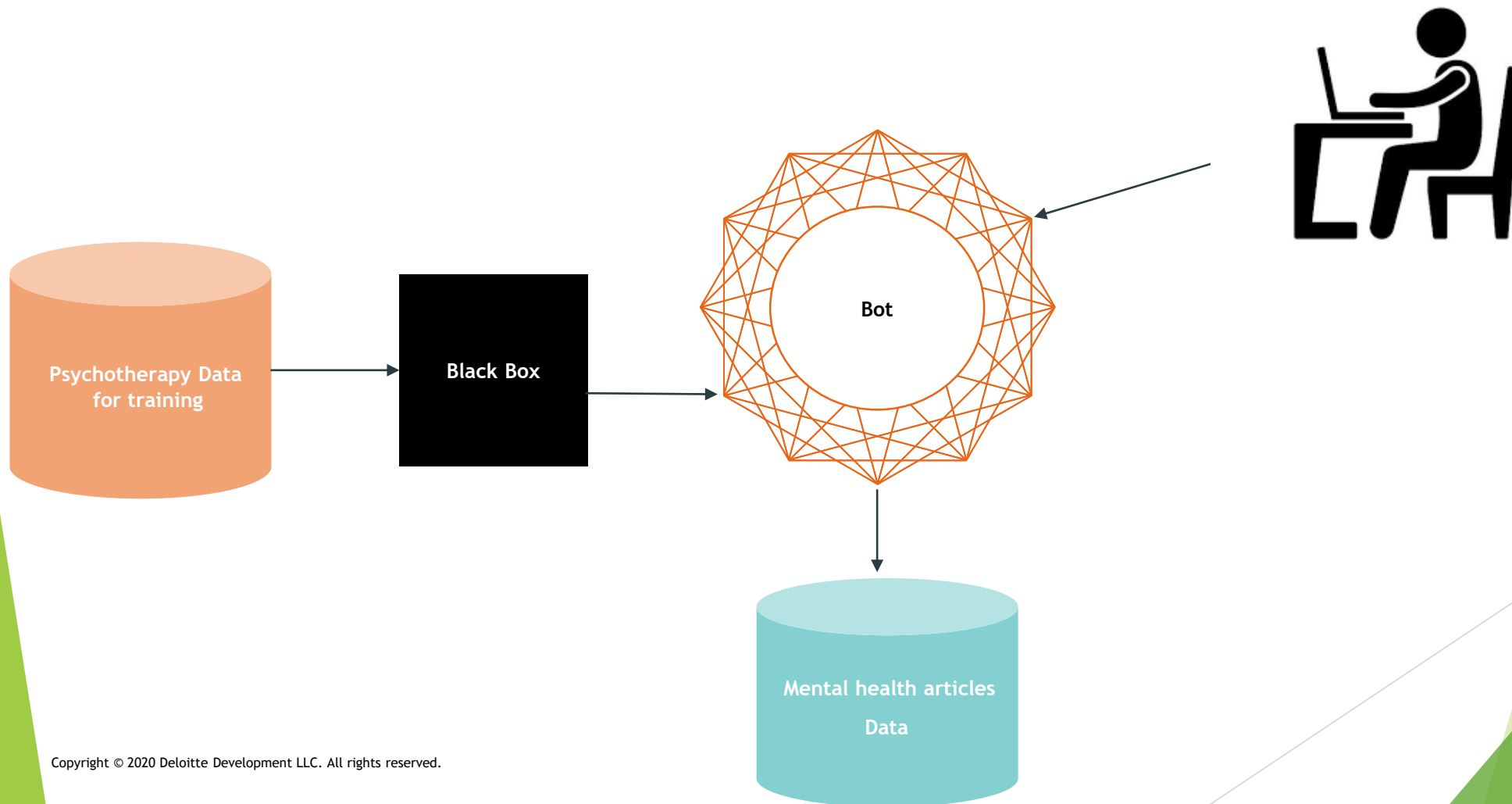


*Right help at the right time can go a long way!*

*Source: National Alliance on Mental Illness*

# Capstone | Overview

If we had a bot friend to check on us on a timely basis, to assess our mood through a chat, help us feel better by sharing some tips on how to deal with a situation or sending along an article for a good read, that can result in not only making one happy, stress free and productive but can also lead to saving their life when suicidal thoughts occur.



# Approach Summary



## Data Sources

[Counselchat.com](https://counselchat.com)  
[7cups.com](https://7cups.com)

Web scraped the above websites related to psychotherapy to extract publicly available questions & their topics.

[Betterhelp.com](https://betterhelp.com)

- The advice page of this website provides articles by topics.
- Once the bot understands the topic to which a user's concern relates to, it would pull out a relevant article from this website and share it with the user.



## Key Fields

'questionText'  
'topic'  
'therapistInfo'  
'therapistURL'  
'answerText'  
'upvotes'  
'views'

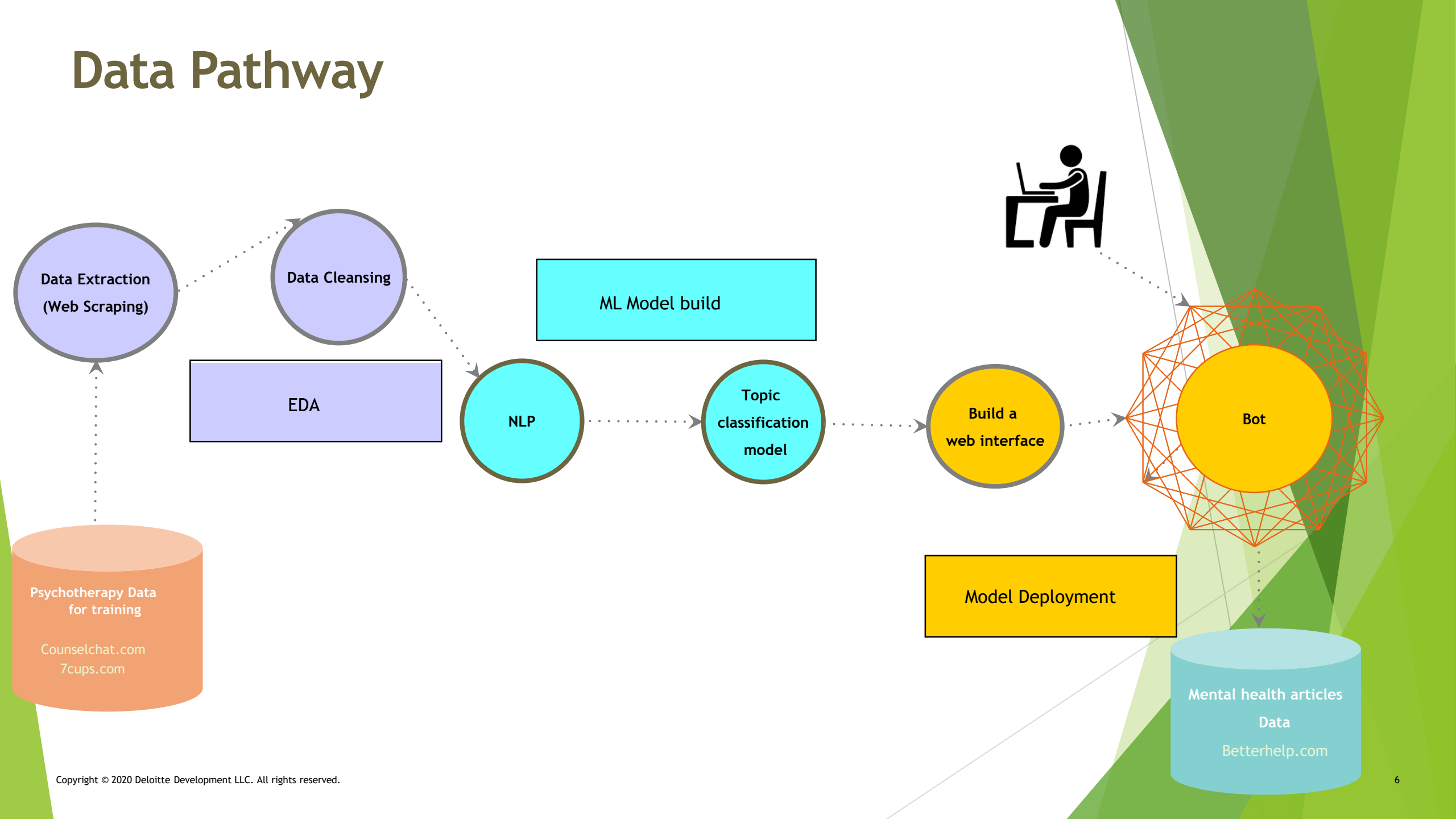
- Every question submitted by the user/patient is classified into a topic/category of mental health.
- Each question can have multiple responses on these sites.
- There is an option to upvote the responses provided to these questions.
- Counselchat.com also shows the number of views for a question.



## Model Results

- Performed EDA on data from the primary source which will be used to train the model.
- EDA helped understand the data and learn answers to questions like:
  - Frequency of questions by topic across different data sources
  - Filter for topics with sufficient data: Anxiety, Depression, Gender issues, Relationship Stress
  - Most popular questions by topic
- Built a classifier to learn the association between the question text and topic with 87.3% accuracy. This classifier will help the bot to sense the topic to which a user's concern relates to.
- Deployed classification model with a Flask based web application.

# Data Pathway



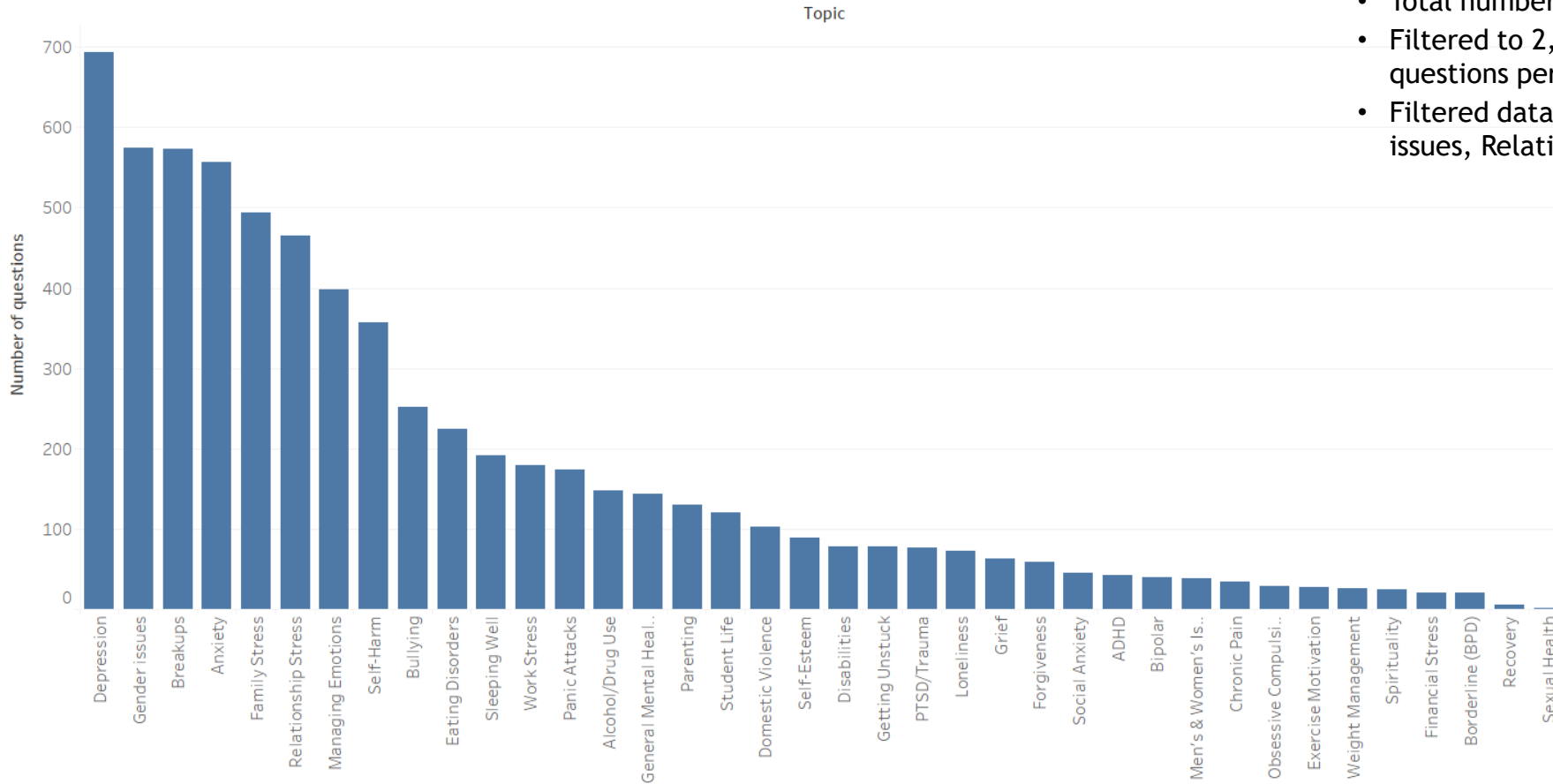
# Exploratory Data Analysis

Data: 2,687 mental health questions & their topics

(Anxiety, Depression, Gender issues, Relationship Stress)



Frequency of questions by topic

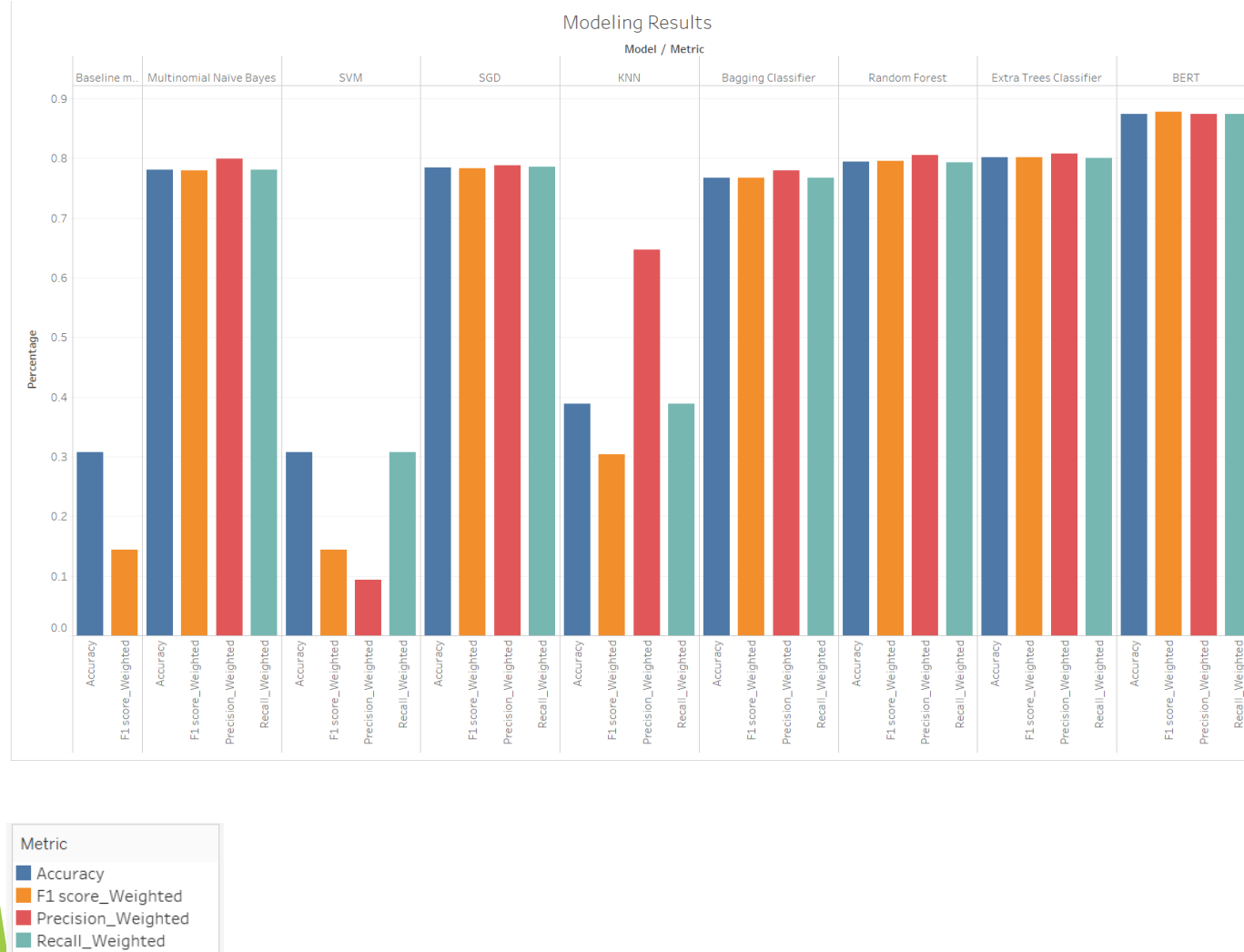


- Total number of records: 6,648
- Filtered to 2,687 records based on the number of questions per topic
- Filtered data contains: Anxiety, Depression, Gender issues, Relationship Stress

# Model Building



Topic classification model: BERT model has provided the highest accuracy of 87.3% & F1 Score\_weighted of 87.8%



## ➤ ML models

- Multinomial Naive Bayes, SGD Classifier, Bagging Classifier, Random Forests, Extra trees classifier.
- Built a baseline dummy classifier which always predicts the most frequent class in the training data.
- Performed Repeated Stratified K-fold Cross Validation for the above models.
- Tree based classifiers performed well against other models with an average accuracy of 80%.

## ➤ BERT (Bidirectional Encoder Representations from Transformers)

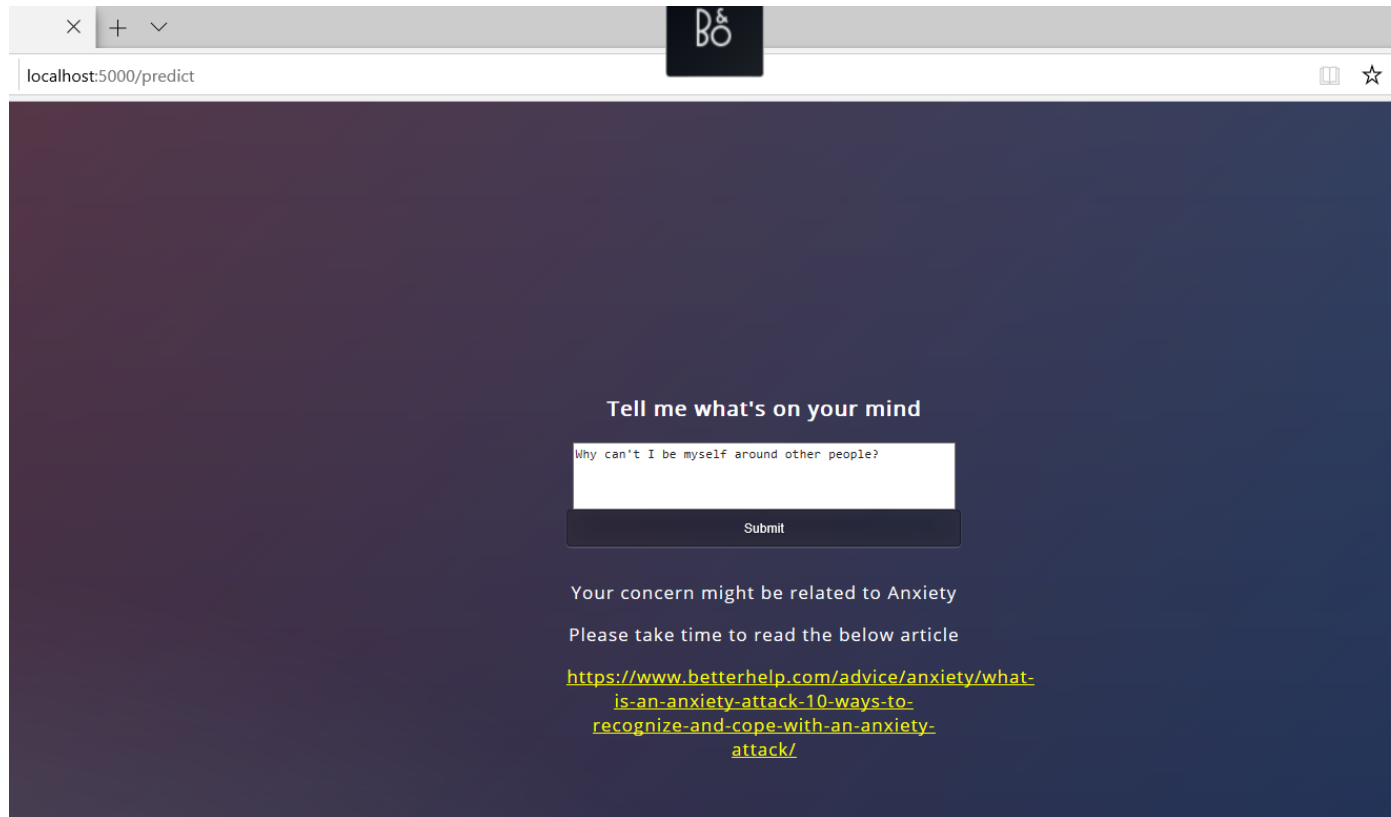
- Used 'Bert For Sequence Classification' module from transformers library.
- Fine tuned the BERT model for classification task.
- Resulted in 87.3% accuracy.



# Deployment

The classifier is made available for the users to interact through an interactive web application. This was done by creating an API for the model using Flask.

## Demo



*Screenshot of web application from Demo*

# Next Steps

## ❑ Gather more data from more topics

- Suicidal ideation data - if available, can help us build a model to create an alert system indicating a need for a professional's involvement

## ❑ Build a good conversational AI assistant & identify more features like the following:

- Can assist the therapist by gathering the pulse & progress of the patient in between sessions.
- Can help the user to be on track with their progress with the help of reminders, exercises/games specific to a user's diagnosis.

## ❑ Use a language model like GPT & GPT-2 for Text generation for the chat bot.

# Summary



Easily Accessible and Available



Efficient assistant, but not a therapist replacement



Aids in CBT (Cognitive Behavioral Therapy) implementation



Build a community of domain experts such as Psychologists, Licensed therapists, Academia etc. as well as more details about the user/patient to personalize it.

# Thank You

# Appendix | References

- <https://www.nami.org/mhstats>
- <https://positivepsychology.com/benefits-of-psychotherapy/>
- <https://machinelearningmastery.com/imbalanced-multiclass-classification-with-the-glass-identification-dataset/>
- <https://towardsdatascience.com/multi-class-text-classification-with-deep-learning-using-bert-b59ca2f5c613>
- <https://medium.com/@aniruddha.choudhury94/part-2-bert-fine-tuning-tutorial-with-pytorch-for-text-classification-on-the-corpus-of-linguistic-18057ce330e1>
- <https://datascience.stackexchange.com/questions/31041/what-does-logits-in-machine-learning-mean>
- <http://davidgratzer.com/reading-of-the-week/reading-of-the-week-can-machine-learning-improve-psychotherapy-the-new-jama-psychiatry-paper-also-santa-ono-on-his-mental-illness/>