

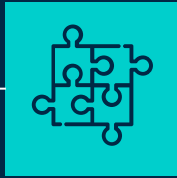


# SUB-REDDIT CLASSIFICATION

Legal & Relationship Advice

Presented By: Peggy Man

# TABLE OF CONTENTS



01

## PROBLEM STATEMENT

Talks about the  
problem and  
challenges



02

## THE PROCESS

Talks about Data  
Cleaning, EDA  
process and  
Modelling



03

## CONCLUSION & RECOMMENDATION

Talks about the  
analysis and  
conclude the  
findings

# PROBLEM STATEMENT

01

# PROBLEM STATEMENT

Will a legal advice be misclassified as relationship advice?

Will a posts be misclassified as relationship issue where it should be label as a legal issue?



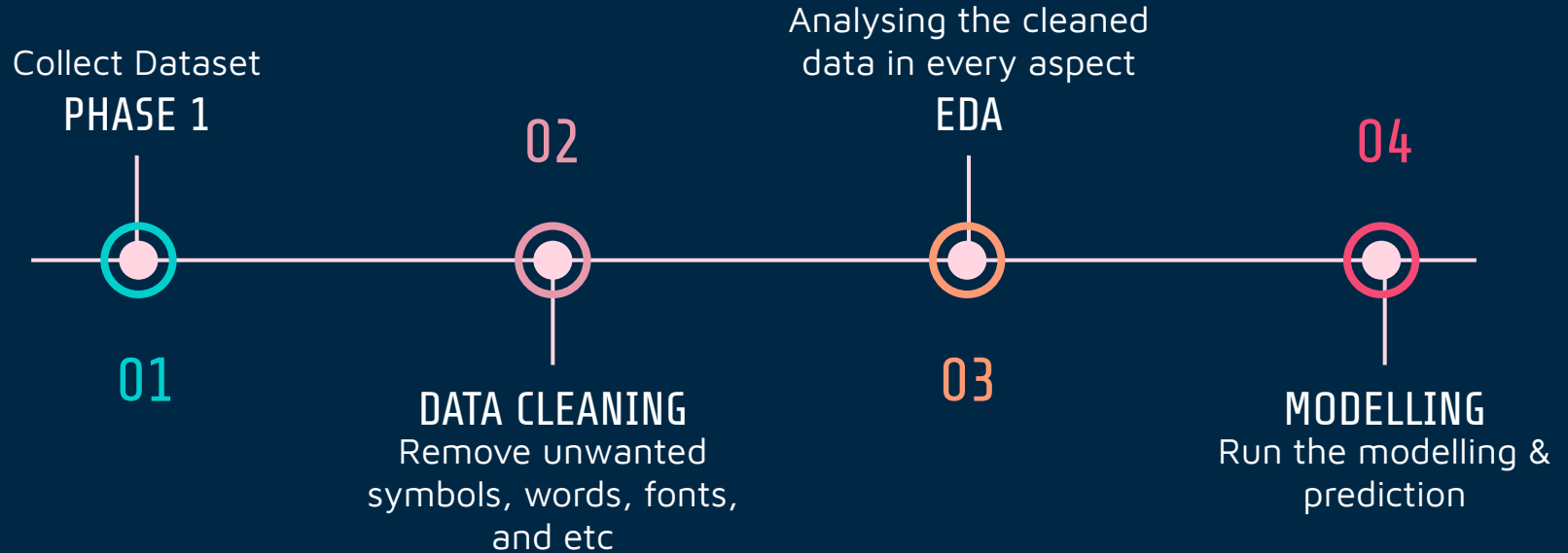
# DATA ANALYSIS



# THE PROCESS

02

# THE PROCESS



# DATASET

REDDIT



POSTS

Legal



1,060

Relationship



1,296



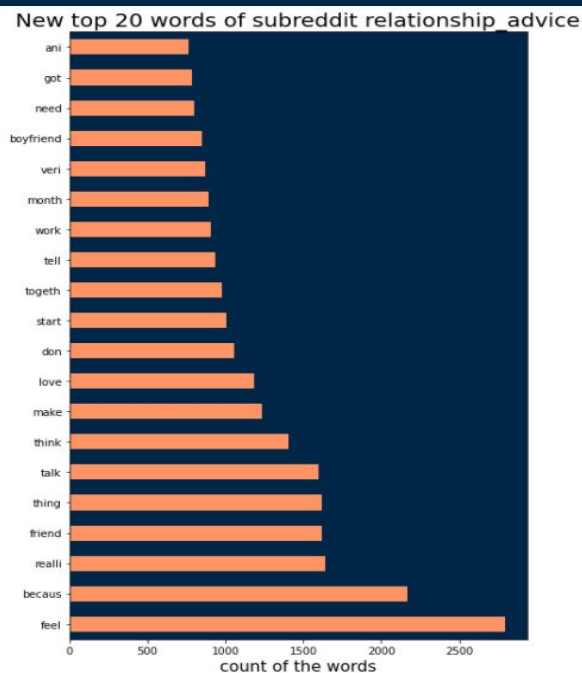
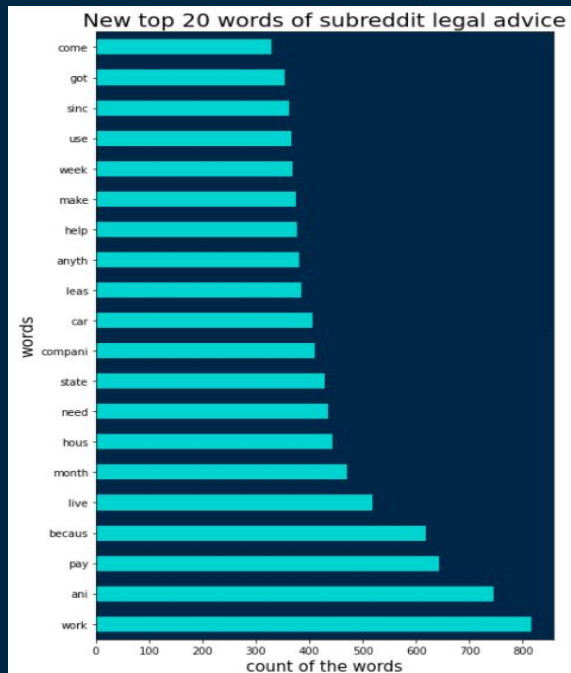
2,356

Total Posts



# EDA IN PROGRESS (Continued.)

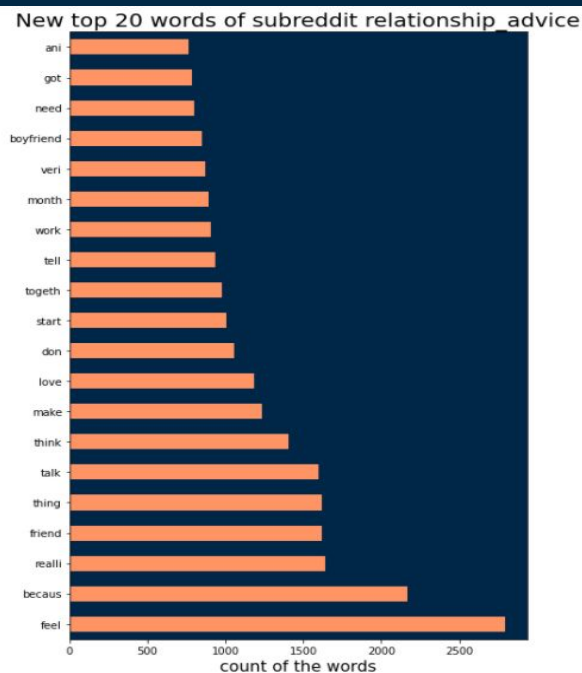
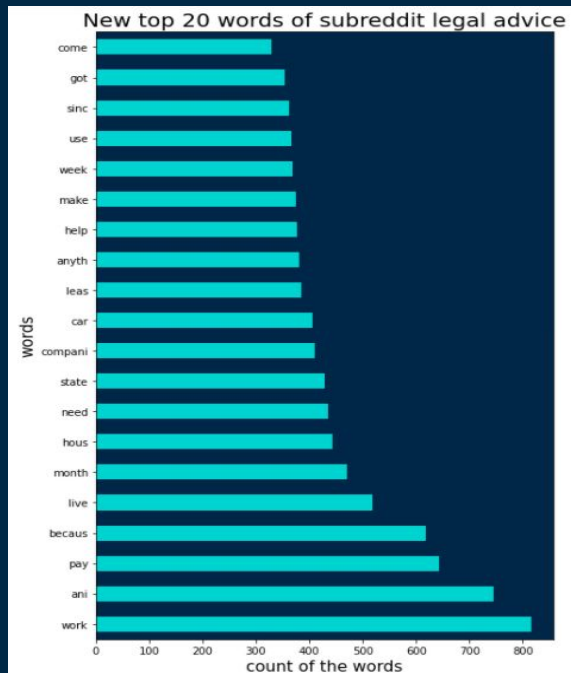
LEGAL



Relationship

# EDA IN PROGRESS (Continued.)

LEGAL



Relationship

# MODELLING

## NAIVE BAYES

Count Vectorizer



Train: 98.5%  
Test: 97.3%

## NAIVE BAYES

TFIDF Vectorizer



Train: 97.8%  
Test: 97.1%

Baseline: 55% of majority classes



## An abstract graphic featuring a dark blue background. It includes a vertical white line on the left, a horizontal white line near the top, a small white square in the upper right, a larger white square in the lower left, a small white square in the center, and a horizontal white line at the bottom right. A solid red rectangle is positioned on the left side, partially overlapping the vertical white line.



## LEGAL ADVICE WORDS CLOUD

## TOP ADVICE



## WORDS CLOUD



# MODELLING (CONTINUED)

## LOGISTIC REGRESSION

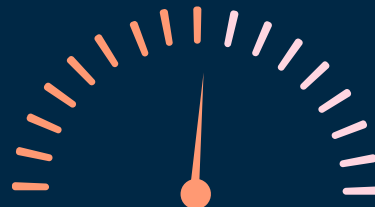
Count Vectorizer



Train: 99.9%  
Test: 97.3%

## LOGISTIC REGRESSION

TFIDF Vectorizer



Train: 96.9%  
Test: 97.4%

Baseline: 55% of majority classes





# CONCLUSION & RECOMMENDATION

03



# ACCURACY:

# 97%



# Analysis of Misclassified Posts

## Findings 1

Examine those legal advice posts that are misclassified

## Findings 2

Mostly related to human relationship. did mention the following words: ex, husband, family, father, mother, stepfather, parents, sexting

## Findings 3

12 misclassified posts. Where it should be legal advice posts but it has been classified as relationship advice posts.

# MISCLASSIFIED POSTS



“Teenager looking for advice on helping a girl...

To start off with why there’s a **criminal tag**, her **stepdad** is a real, and excuse my language, but the human shitstain makes me sick whenever I think about what he’s done. He has to list: **raped** her at least twice, **beaten** er, tried to touch her before, makes her feel like she can’t breathe”

“I’m begging you, **Reddit**, help me find out what she could do to get out of that situation, or if any **custody lawyers** read this, tell me where she’ll go.”



“Back to the list of **offenses** by fuckstick of a **stepdad**: pulled a **gun** on her when he was **drunk**, **threatened to kill** her when she tried to run away, and undoubtedly more that she hasn’t told me. I’ve advised **family**, but they’re all in different states.”

—Classified as Relationship Advice but the post is from Legal Advice

# MISCLASSIFIED POSTS



"Scared, **paranoid** of sexting mistake I'll keep this. It's not a **scam**"

**Note:** Sexting is sending, receiving or forwarding sexually explicit messages, photographs, or images, primarily between mobile phones, of oneself to others.

"**I don't know if this the right subreddit** but I just want to reach as many people as possible to find a solution. I'm 21 and I live at home with my **mother** who frequently gets **beaten** by my **stepfather**."



"So I don't know how to tag this but my **ex** all the sudden out of the blue contacts me and is **threatening** to take my cat throw away. What do I do if she comes to my **parents** house searching for me or the cat despite not being welcome there?"

—Classified as Relationship Advice but the post is from Legal Advice

“Collect at least 10,000 posts and to include other features like the comments, created time, length of post, run on others classification models such as Random Forest and K-Nearest Neighbour. Try different stop words with different parameters of the models.”

—IMPROVEMENT PLAN

“Educate Reddit admins and the Reddit moderators to be **more aware** and **pay attention** to cases involving family and parenting problems”

“Can be **critical** and **sensitive**, we should **minimize** the misclassified post.  
(False negative are more important than the false positive.)”

—RECOMMENDATION

Do you have any questions?

# THANKS



CREDITS: This presentation template was created by [Slidesgo](#),  
including icons by [Flaticon](#), and infographics & images by [Freepik](#)  
Please keep this slide for attribution