



**MSF** | MINISTRY OF  
SOCIAL AND FAMILY  
DEVELOPMENT

## Stakeholder Meeting

# Persona Introduction

# Persona Introduction



## George, 33

Campaign Strategist

- Assigned a project to create campaigns for mothers.
- Had troubles classifying articles or post written by mothers.
- Noticed that he and his team had discrepancies in the findings, complicating the analysis process.



# Mothers

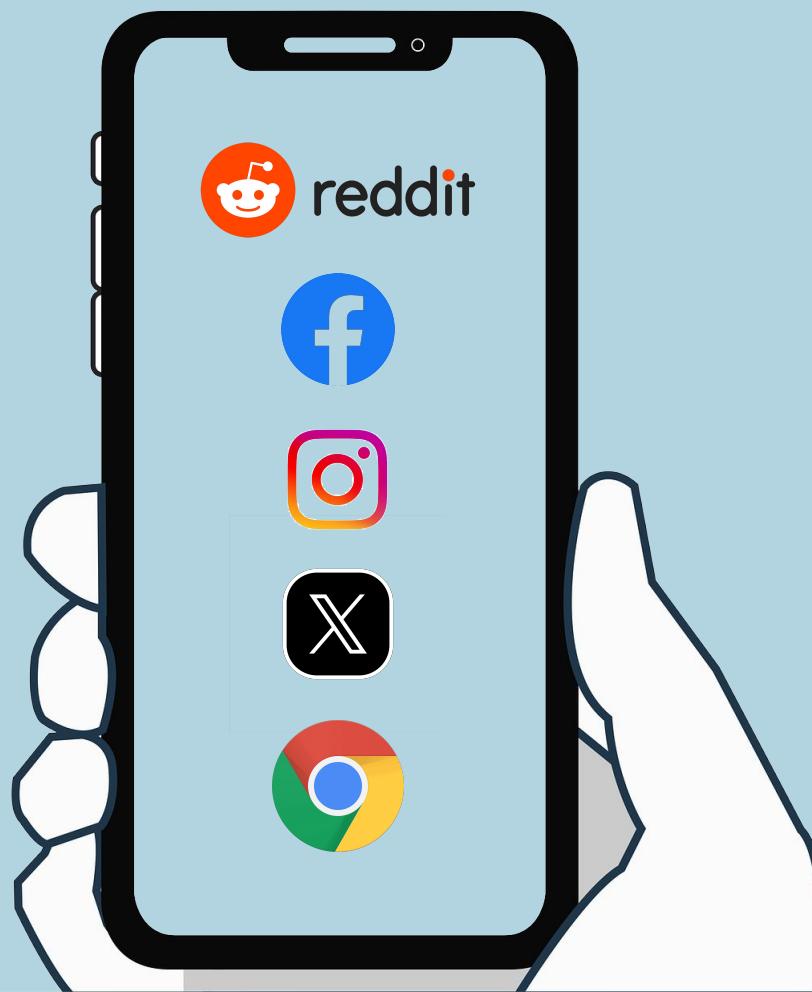
# Mothers

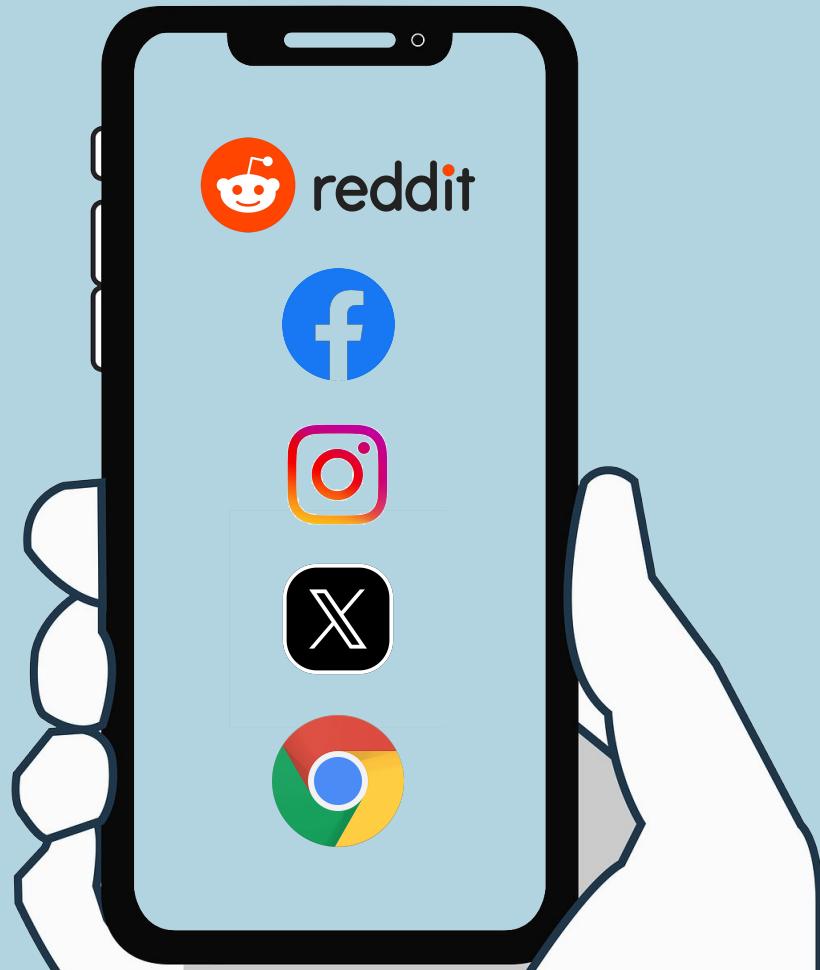
Work life



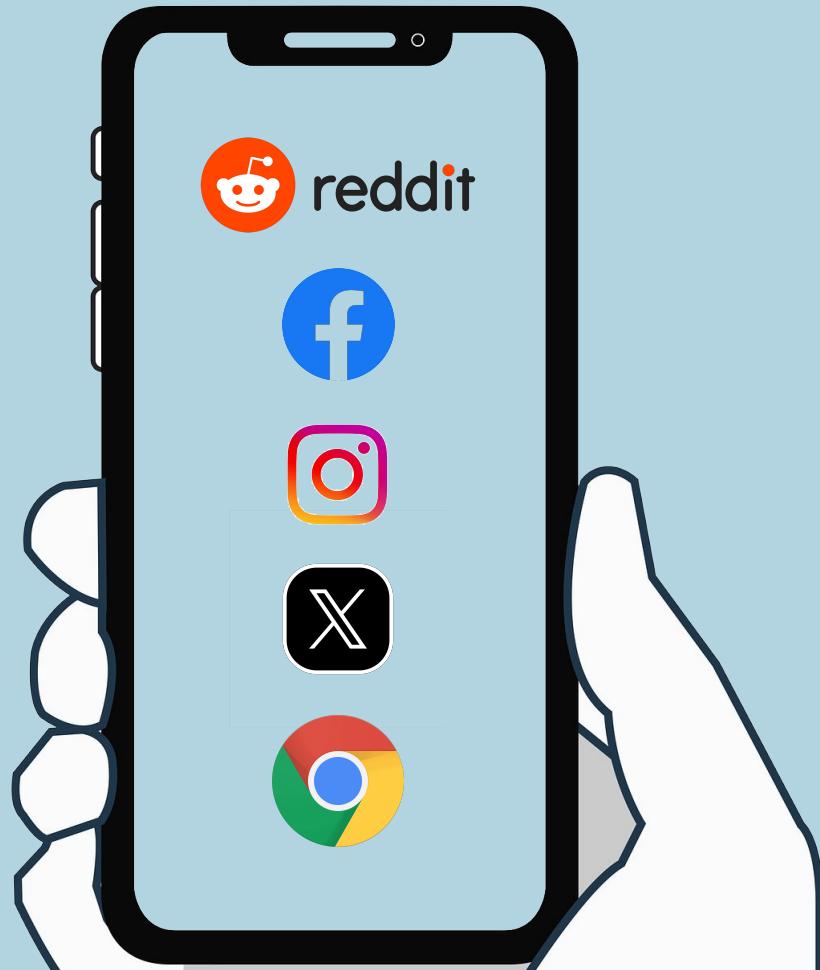
Parenting



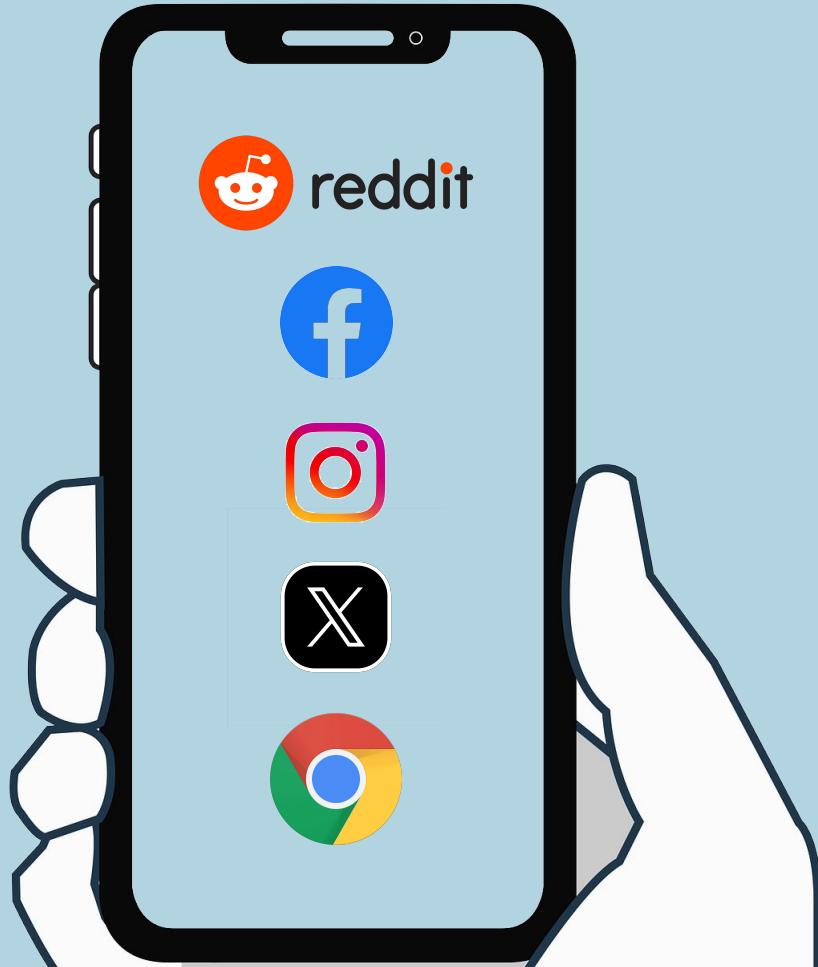




# Information



# Information Discussions



Information  
Discussions  
Anonymous  
Support

 u/Sush1burrito · 2 hr. ago

## Should I move in with my dad? I'm a s...

My baby is 5 months old. You've probably seen my apartment, low income apt, and it's amazing. I love it! I have tons of room, and the location is awesome.

## Prenatal Yoga Recommendations in Jurong West

 Jess\_SeptemberMum · Feb 20, 2024



Feb 20, 2024

Hi,

Searching for Prenatal yoga recom...

Saw in previous threads from 2015

Much thanks in advance!

## Confinement nanny to recommend

 e\_Kaix · Dec 6, 2023



Dec 6, 2023

I would like to recommend Auntie Annie for those who are looking for a confinement nanny.

She constantly reminded me that I have to get ample of rest while I can count on her to take care of the baby. Baby has jaundice and she ensured that baby was fed every 3 hours so we can clear the jaundice asap. As a new mother, I believe there are many panic moments which we find ourselves lost and panicking. With an experienced nanny like Auntie Annie, I can trust that the baby is well taken care of.

 u/DCDPTinCP · 4 hr. ago



es for the meals she cooked. She made use of the differences. Husband is

... period while I pick things

## Daycare moms, how are you surviving illnesses?

My son started daycare at 12 weeks. Since then, he's been on and off sick. Sleep is a mess, etc. But the real problem is that I am also constantly sick. I've had strep twice, 3 ear infections, and various colds since August. I currently have strep again and haven't been able to eat since Friday. I'm lucky my husband is a saint and takes over as much as possible, but I am struggling with feeling useless. How are you all surviving? Is there a trick I am missing? Before baby was born I hadn't been sick since 2019 - thank you handwashing and social distancing. Before that, I'd get sick MAYBE twice a year. I pride myself in exercising and eating well, but that is no match fo...

### Similar threads

 6 

 16

 Share

 Measuring 2 weeks behind  
Vanella · Dec 22, 2023 · Year 2024 Mums

Replies: 6  
Views: 1K  
Dec 28, 2023  Vanella

 Singapore should beware of the pitfalls of co-funding private IVF treatment  
Angelica Cheng · Feb 29, 2024 · Trying To Conceive

Replies: 2  
Views: 971  
Mar 5, 2024  Angelica Cheng

 Unable to conceive for almost 1 year. HSG/Doctor recommendation?  
mangosaladbk · May 29, 2022 · Trying To Conceive

Replies: 0  
Views: 4K  
May 29, 2022  mangosaladbk

 Husband backed out on having kids  
sw08x · Feb 6, 2024 · Motherhood, Pregnancies And Babies

Replies: 3  
Views: 1K  
Feb 8, 2024  Positivetots

 Is it worthwhile doing IVF genetic testing (PGS / PGT-A) to prevent birth defects in older women?  
Angelica Cheng · Oct 7, 2021 · Trying To Conceive

Replies: 0  
Views: 4K  
Oct 7, 2021  Angelica Cheng

**Singapore]**

# Problem Statement



In light of the current generation of mothers' high engagement on social media and forums, particularly in discussions related to work-life balance and childcare, our Campaign Strategists face challenges in consistently identifying and classifying relevant topics for our campaigns.

This reliance on personal assumptions rather than empirical evidence hampers the effectiveness of our campaigns or programs, hindering our ability to optimize online presence and effectively reach our target audience.



# Classification Model

---

# Classification

Mothers

Fathers



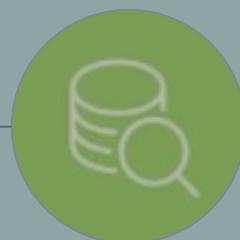
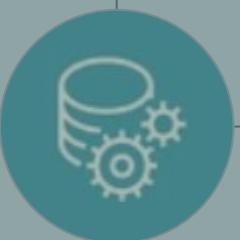
Define the problem



Data Collection



Data Cleaning & Preprocessing



Exploratory Data Analysis

Modelling



Model Evaluation

# Criterias for selection

---



Relevance



Activity &  
Engagement level



Membership  
Size



Moderation



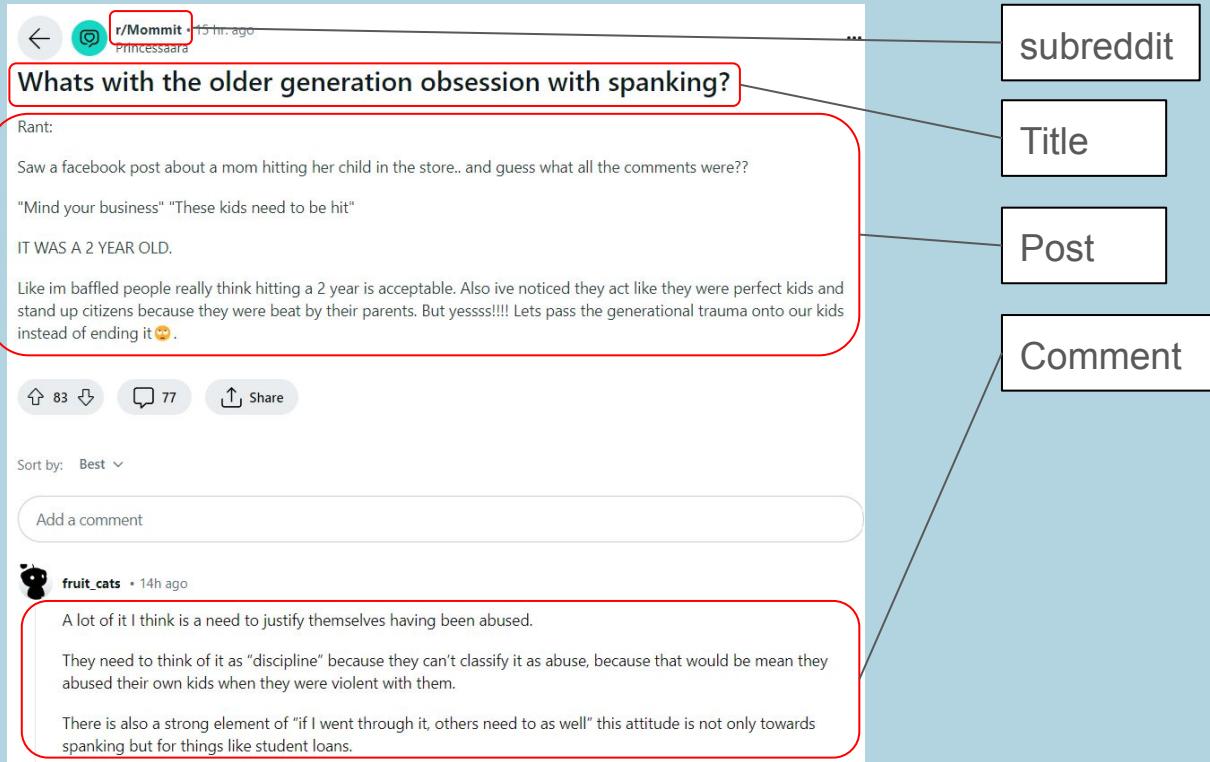
Quality of Content



Variety



# Text scraping





Define the problem



Data Collection



Data Cleaning & Preprocessing



Exploratory Data Analysis



Modelling



Model Evaluation

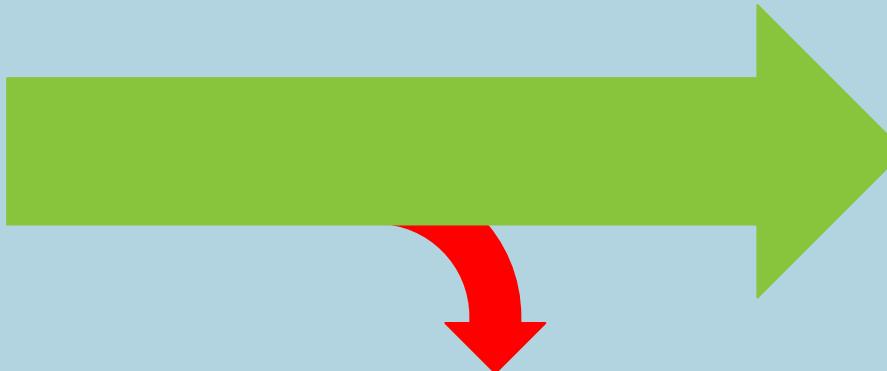


# Data Cleaning

---



Raw data collected



Cleaned data

<praw.models.comment\_forest.CommentForest object at 0x31f8ec310>

. , + , \* , ? , ^ , \$ , ( , ) , [ , ] , { , } , | , \ , ! , @ , # , % , &

Blank rows

Auto-replies or system messages

[removed] and [deleted]

\n

# Preprocessing

"this was my daughter we tried everything and the only thing that worked was being held i do not have any tips because nothing worked for us when she was around 18-20 months i coslept with her because i was falling apart from sleep deprivation she is 3 years old now and sleeps through the night in her own bed with zero issues so while it has been an absolute brutal couple of years shes a great independent sleeper now"

## Tokenization

1

```
['this', 'was', 'my', 'daughter', 'we', 'tried', 'everything', 'and', 'the', 'only', 'thing', 'that', 'worked', 'was', 'being', 'held', 'i', 'do', 'not', 'have', 'any', 'tips', 'because', 'nothing', 'worked', 'for', 'us', 'when', 'she', 'was', 'around', '18-20', 'months', 'i', 'coslept', 'with', 'her', 'because', 'i', 'was', 'falling', 'apart', 'from', 'sleep', 'deprivation', 'she', 'is', '3', 'years', 'old', 'now', 'and', 'sleeps', 'through', 'the', 'night', 'in', 'her', 'own', 'bed', 'with', 'zero', 'issues', 'so', 'while', 'it', 'has', 'been', 'an', 'absolute', 'brutal', 'couple', 'of', 'years', 'shes', 'a', 'great', 'independent', 'sleeper', 'now']
```

## Stop Word removal

2

```
['daughter', 'tried', 'everything', 'thing', 'worked', 'held', 'tips', 'worked', 'us', 'around', '18-20', 'months', 'coslept', 'falling', 'apart', 'sleep', 'deprivation', '3', 'years', 'old', 'sleeps', 'night', 'bed', 'zero', 'issues', 'absolute', 'brutal', 'couple', 'years', 'shes', 'great', 'independent', 'sleeper']
```

## Lemmatization

3

```
['daughter', 'try', 'everything', 'thing', 'work', 'hold', 'tip', 'work', 'us', 'around', '18-20', 'month', 'coslept', 'fall', 'apart', 'sleep', 'deprivation', '3', 'year', 'old', 'sleep', 'night', 'bed', 'zero', 'issue', 'absolute', 'brutal', 'couple', 'year', 'she', 'great', 'independent', 'sleeper']
```



Ready for Exploratory Data Analysis



Define the problem



Data Collection

Data Cleaning & Preprocessing



Exploratory Data Analysis

Modelling

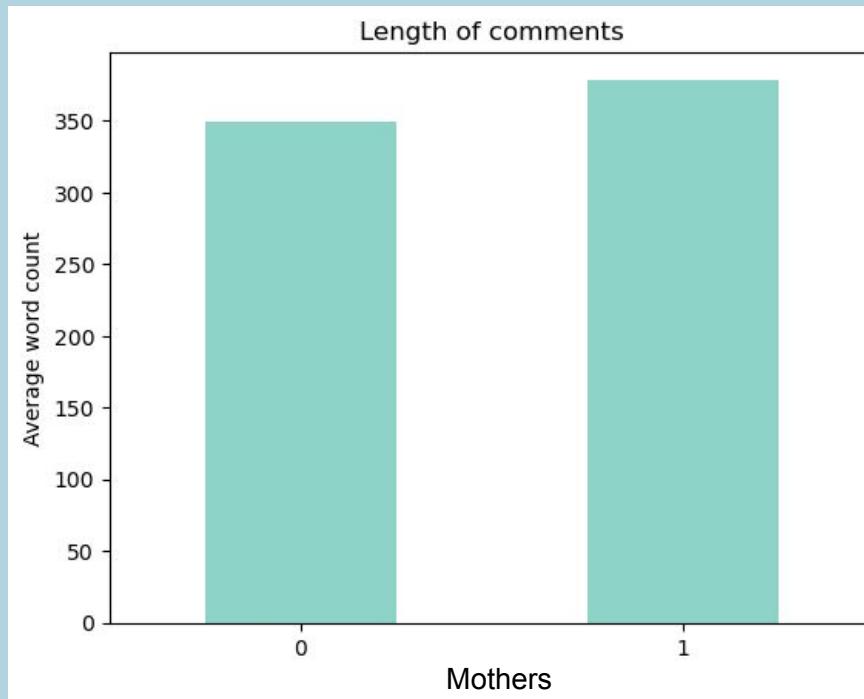


Model Evaluation

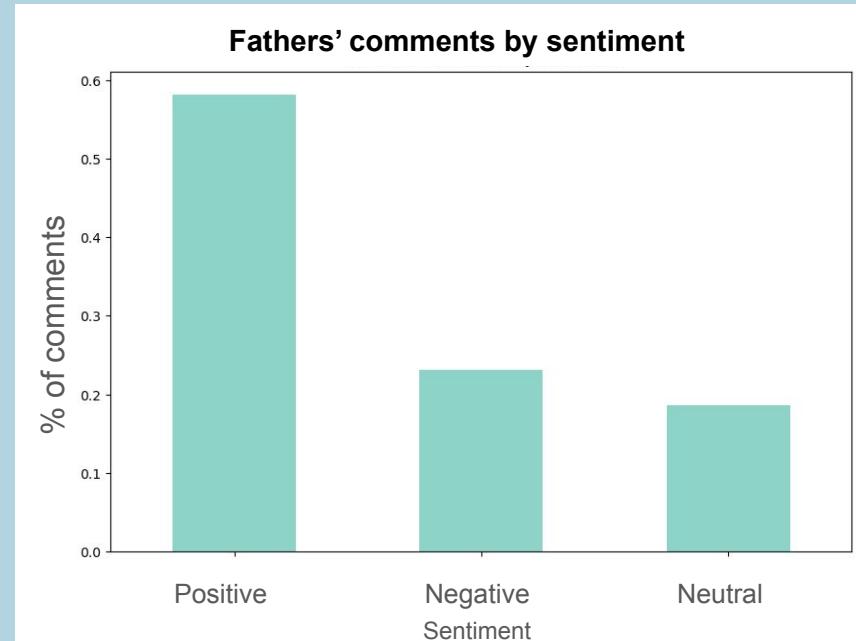
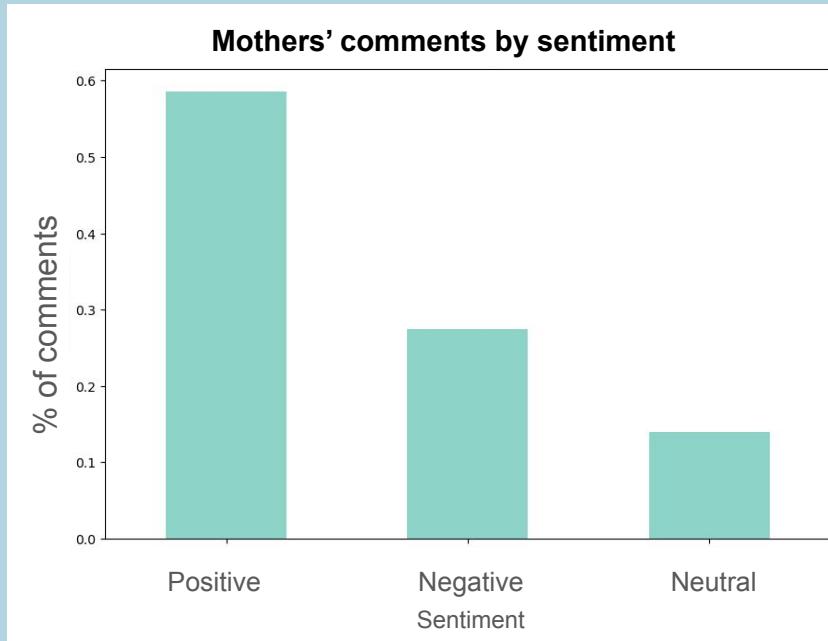
# Exploratory Data Analysis

- 1) Comment and post length
- 2) Sentiment Analysis
- 3) N-gram analysis

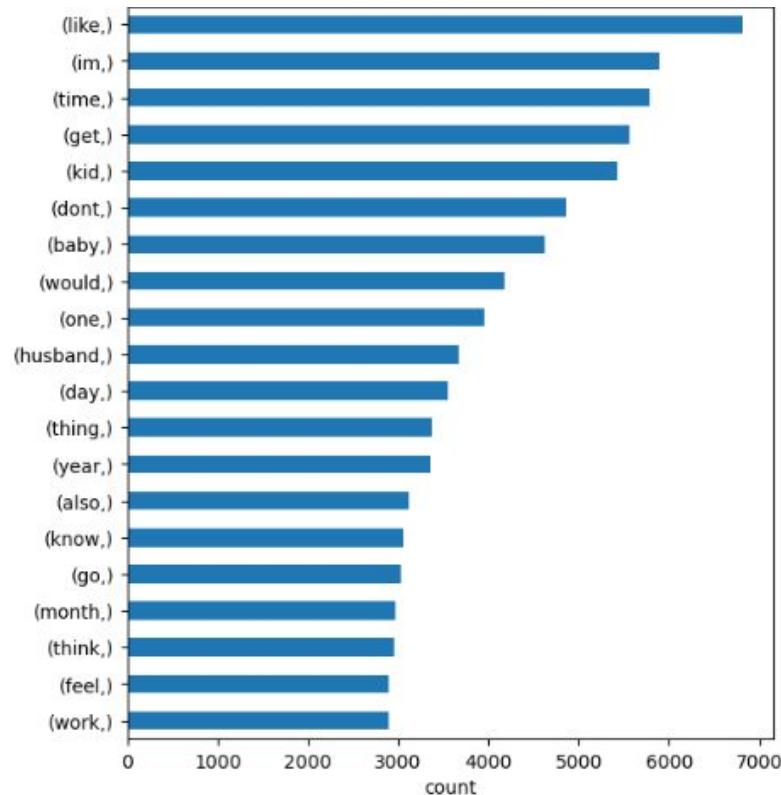
# 1) Length of comments



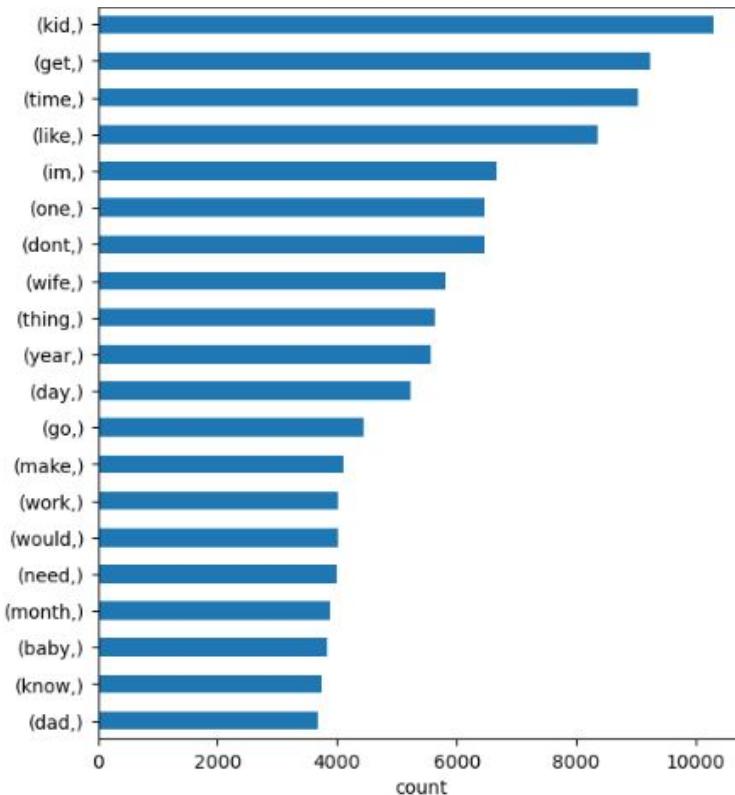
## 2) Sentiment Analysis



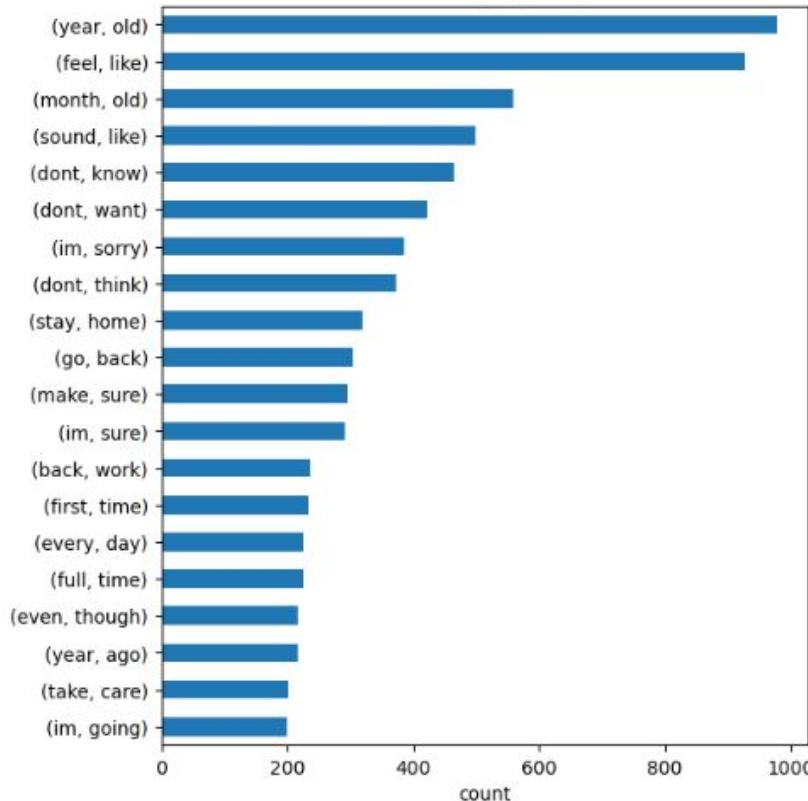
### Mummit (1-gram)



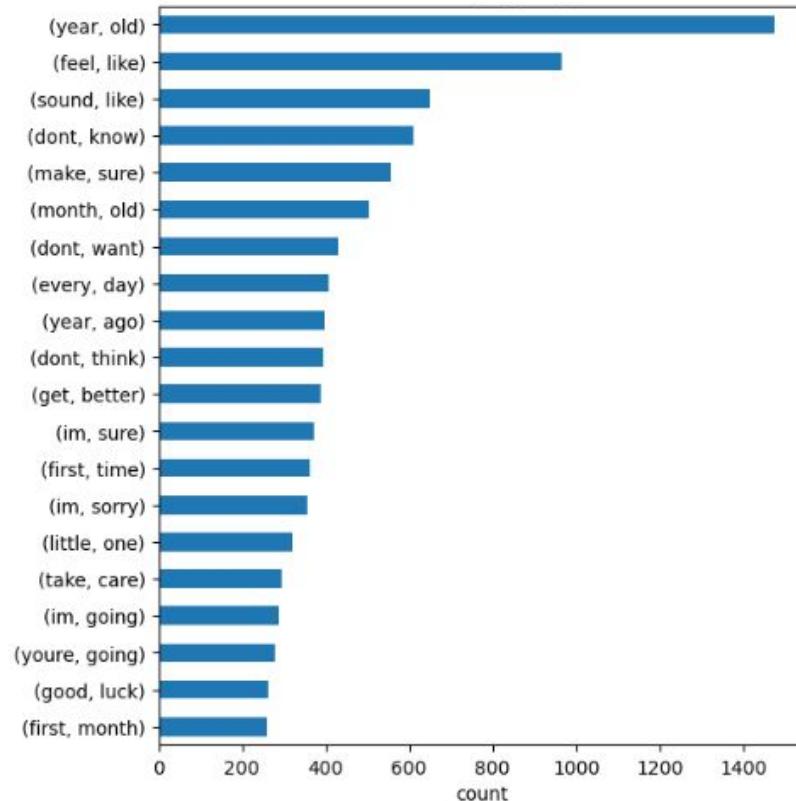
### Daddit (1-gram)



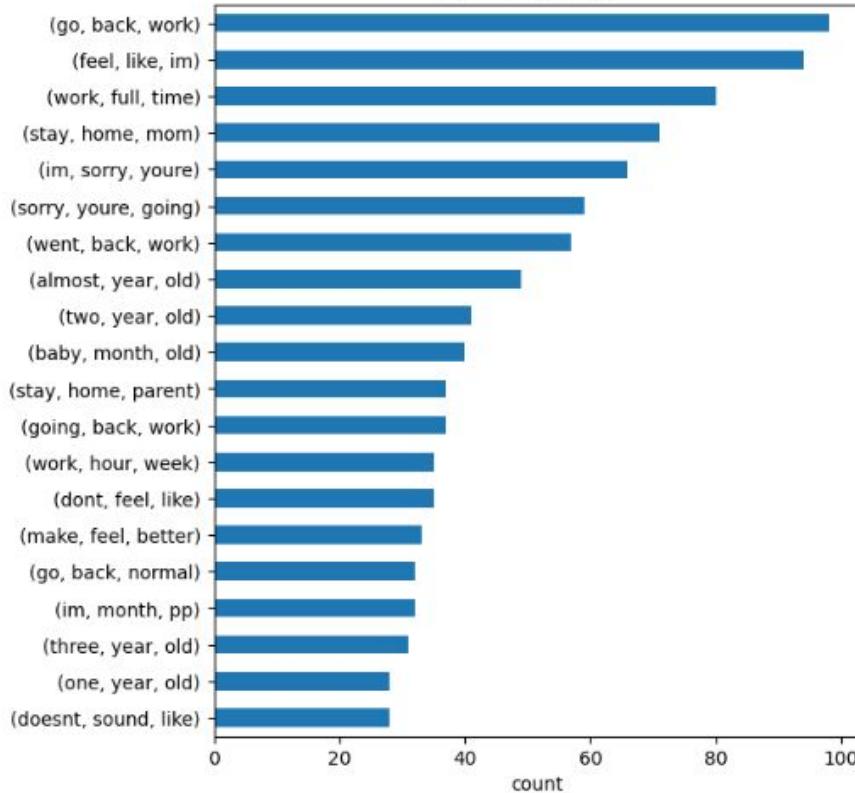
### Mummit (2-gram)



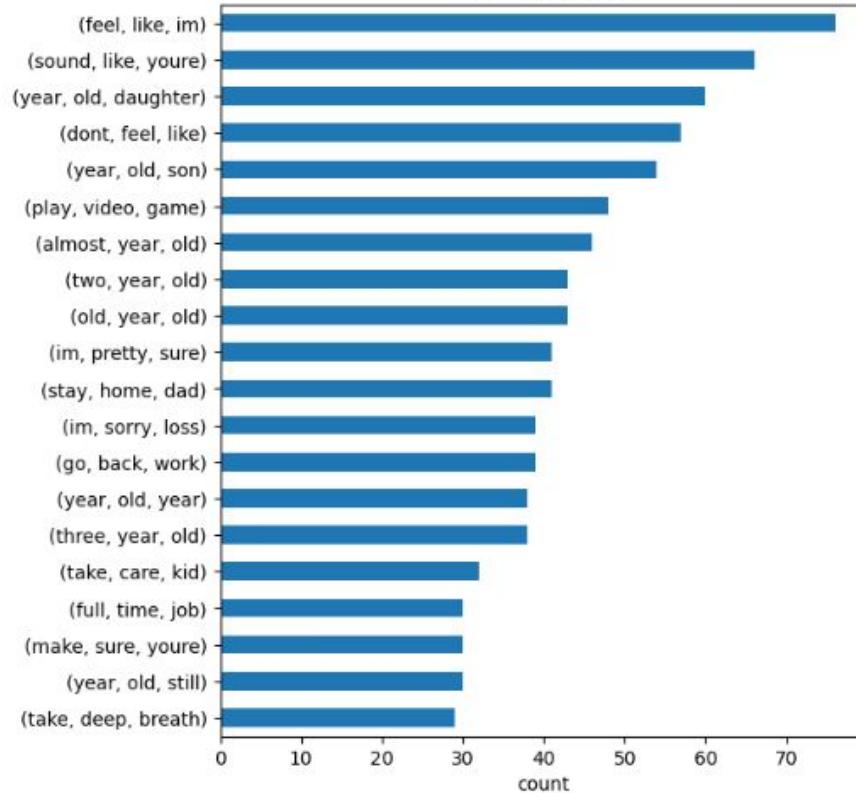
### Daddit (2-gram)



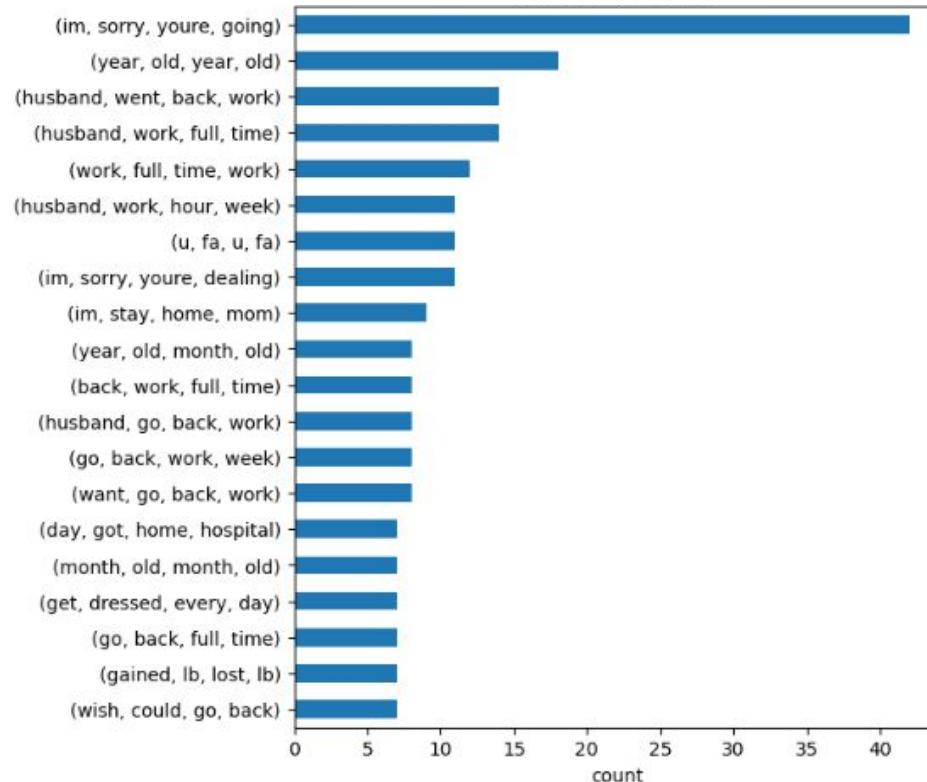
### Mummit (3-gram)



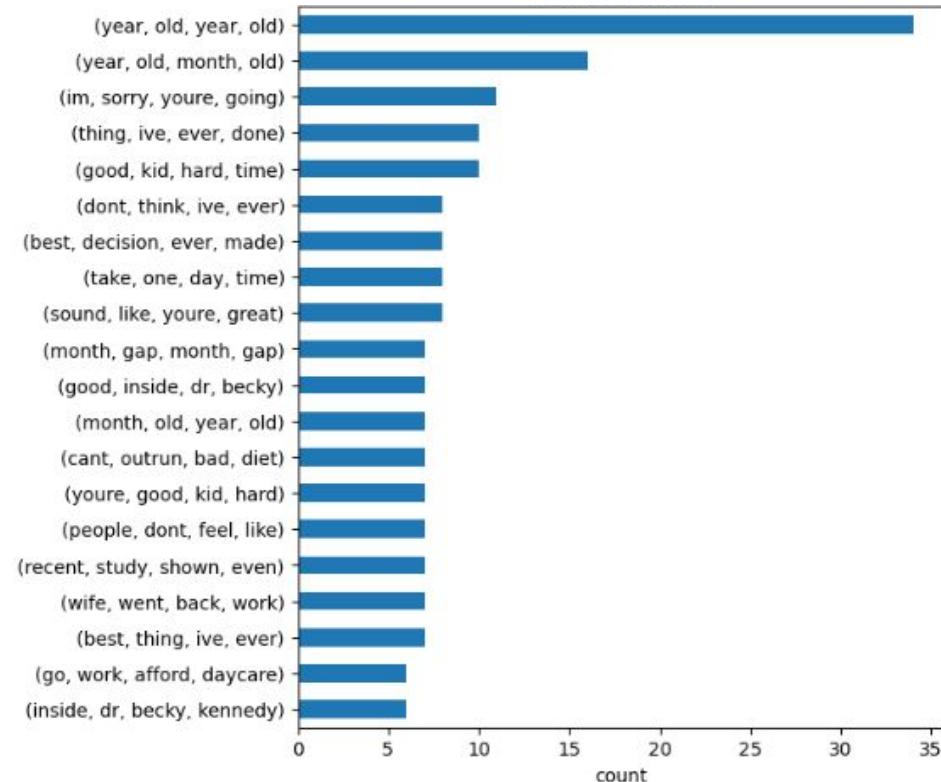
### Daddit (3-gram)



### Mummit (4-gram)



### Daddit (4-gram)





Define the problem



Data Collection

Data Cleaning & Preprocessing



Exploratory Data Analysis

Modelling



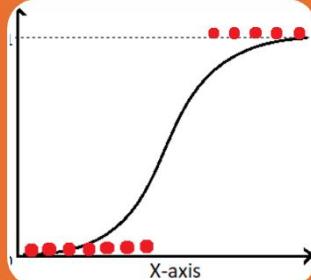
Model Evaluation

# The Model

---

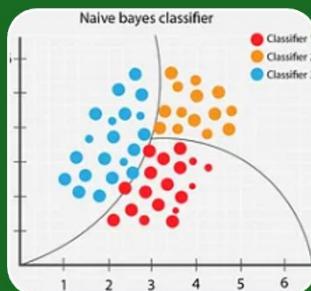
# Types of models

---



## Logistic Regression

- Predicts the probability of binary outcome.
- Handles various data types and provides insights into feature importance.



## Multinomial Naïve Bayes

- Classifies text-like data where features represent counts (e.g., word frequencies).
- Fast and simple, based on Bayes Theorem for binary data

# Types of text vectorizers

## Count Vectorizer

- Converts text documents into a matrix where each element counts the occurrences of a word in a document.

## TF-IDF vectorizer

- Similar to CountVectorizer, but it weights words based on their frequency in a document and rarity across documents.



Define the problem



Data Collection

Data Cleaning & Preprocessing



Exploratory Data Analysis

Modelling



Model Evaluation

# Model scores

---

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798
2	Naive Bayes - Multinomial	TF-IDF Vectorizer	0.700	0.640	0.514	0.848	0.942	0.776
3	Logistics Regression	CountVectorizer	0.845	0.710	0.656	0.775	0.879	0.792
4	Logistics Regression	TF-IDF Vectorizer	0.780	0.716	0.650	0.797	0.895	0.800

# Model scores

---

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798
2	Naive Bayes - Multinomial	TF-IDF Vectorizer	0.700	0.640	0.514	0.848	0.942	0.776
3	Logistics Regression	CountVectorizer	0.845	0.710	0.656	0.775	0.879	0.792
4	Logistics Regression	TF-IDF Vectorizer	0.780	0.716	0.650	0.797	0.895	0.800

F1 score is our primary metric.

# Metrics

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798

## F1 score:

Harmonic mean of sensitivity and precision, combining both metrics into a single score that penalizes models with uneven class distribution.

Best balance of precision and sensitivity

# Metrics

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798

Sensitivity:

The proportion of actual positive cases that the model correctly identified (true positives out of all positive examples).

For example: 
$$\frac{\text{Correct prediction "mum"}}{\text{Total "mum" comments}} = \frac{70.9}{100} = 0.709$$

# Metrics

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798

Precision:

Measures the proportion of positive predictions that were actually correct (positive predictive value).

For example: 
$$\frac{\text{Correct prediction "mum"}}{\text{Total prediction "mum}} = \frac{75.5}{100} = 0.755$$

# Metrics

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798

Precision:

$$\text{Sensitivity} = \frac{\text{Correct prediction "mum"}}{\text{Total "mum" comments}}$$

Measures the proportion of positive predictions that were actually correct (positive predictive value).

For example:  $\frac{\text{Correct prediction "mum"}}{\text{Total prediction "mum"}} = \frac{75.5}{100} = 0.755$

# Why F1 score?

	Model	Vectorizer	Precision train data	Precision on test data	Sensitivity	F1 score	Accuracy
	Logistics Regression	TF-IDF vectorizer	0.976	0.933			

Initially we used **Precision** as our main metric to assess the model.

The scores were very good.



# Why F1 score?

---

	Model	Vectorizer	Precision train data	Precision on test data	Sensitivity	F1 score	Accuracy
	Logistics Regression	TF-IDF vectorizer	0.976	0.933	0.001	0.023	0.615

But sensitivity was very low!

# Model scores

---

	Model	Vectorizer	F1 score on train data	F1 score on test data	Sensitivity	Precision	Specificity	Accuracy
1	Naive Bayes - Multinomial	CountVectorizer	0.765	0.731	0.709	0.755	0.854	0.798
2	Naive Bayes - Multinomial	TF-IDF Vectorizer	0.700	0.640	0.514	0.848	0.942	0.776
3	Logistics Regression	CountVectorizer	0.845	0.710	0.656	0.775	0.879	0.792
4	Logistics Regression	TF-IDF Vectorizer	0.780	0.716	0.650	0.797	0.895	0.800

Final model = Naive Bayes with CountVectorizer



# **Limitations of the Model**

---

# Limitations of the Model

## 1) **Unable to identify localized words**

The model was trained on data from the global subreddits r/daddit and r/Mummit

## 2) **Performance of the model is constrained by limitations in the selected models.**

Exhaustive tuning did not yield substantial improvements.



# Classification

Mothers

Fathers

# How can we apply this Classifier model?

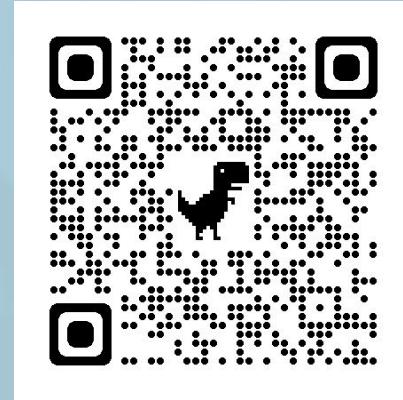
---

# Afternoon Tea

Finding Top Words Mentioned From Mothers

Afternoon tea  
App

# Afternoon Tea App



# Demo



# George, 33

Campaign Strategist

- ✓ Had troubles classifying articles or post written by mothers.
- ✓ Noticed that he and his team had discrepancies in the findings, complicating the analysis process.

# What's Next

---

## With more resources:

- We can acquire more localized data from either from other social media platforms or forums.
  - For example: Picking up Singlish.
- Use more complex models such as Random Forest or XGBoost.
  - To have more accurate classification.

# Thank you!

---

