The purpose of this deck is to fulfill the requirement of DSI 39, Project 3.

Role: Ministry of Social and Family Development (MSF) Data Science Team

Audience: MSF's Senior Management

Problem Statement: Declining number of marriages and earlier divorces are a cause for concern. With one of the key reasons for divorces being communication problems, how can we help couples facing difficulties in their relationship, strengthen communication, through recognizing and understanding their attachment styles?

Done by: The Colony (Jackie Seah, Li Cheng, Mei Qi, Yvonne Lim)

A self-help tool for couples to understand their attachment style

MSF Senior Management Meeting
29 Sep 2023

Meeting Agenda

1. Marriage landscape in Singapore
2. Methodology
3. Summary of results
4. Prototype of our solution

THE STRAITS TIMES

Number of marriages registered continues to decline, more couples divorced in 2019

Divorces in S'pore rise by 2% in 2021

More S'pore couples, especially those recently married, getting divorced: MSF

Recent trends in declining number of marriages and earlier divorces are a cause for concern

↓4.6% **Annual avg marriage registered**
(comparison between 2018-2022 and 2013-2017)

1.9 **Avg crude divorce rate in the last 10 years**
(per 1000 resident)

$10,000-$35,000 **Cost for contested divorce**

Notes:

- *Data on marriage and divorce from Department of Statistics Singapore*
- *Crude divorce rate refers to the number of divorces and annulments granted in the year per thousand population.*
- *Data on cost of divorce from DBS*

Communication is one common reason for divorce.

A strong and lasting relationship requires healthy communications.

Attachment styles influence how we connect with others

- Comfortable communicating needs and with closeness

Secure

Anxious

- Feel compelled to maintain frequent communication due to constant need of approval and reassurance
- Tend to overanalyze partner's message and behaviour due to fear of rejection

- Self-reliant
- Hesitant to express actual emotions
- Difficulty in verbal expression of love

Avoidant

Fearful

- Inconsistent behaviour due to internal conflict between desire for intimacy and fear of it

How can we help couples
facing difficulties in their relationship,
strengthen communication,
through recognising and understanding
their attachment style?

To develop a **self-help tool** that users can use easily to **identify attachment style of both** themselves and their partner's, with an **accuracy of more than 0.8**

Work Flow

01

Procuring Data Source

Scraping post data from 2 subreddits to use as labelled data for the machine learning model.

02

Data Cleaning and Analysis

Cleaning and processing the dataset to sift out the distinct features.

03

Modelling

Training and evaluating multiple machine learning models to be able to make accurate predictions.

04

App Design

Designing an web application that is user-friendly for couples to self-help.

Labelled data is sourced from Reddit to feed in ML model

Reddit is a website where registered users (or "redditors") post content. Posts are organized by subject into user-created boards called "communities" or "subreddits".

There are subreddits where people with specific attachment styles gather to share experiences and provide support for each other.

- r/AnxiousAttachment. 38k members
- r/AvoidantAttachment. 19k members.

Posts scrapped from these 2 subreddits are then used as labelled data to feed into a machine learning model. The assumption is that posts made in these subreddits reflect the patterns of someone with said attachment style.

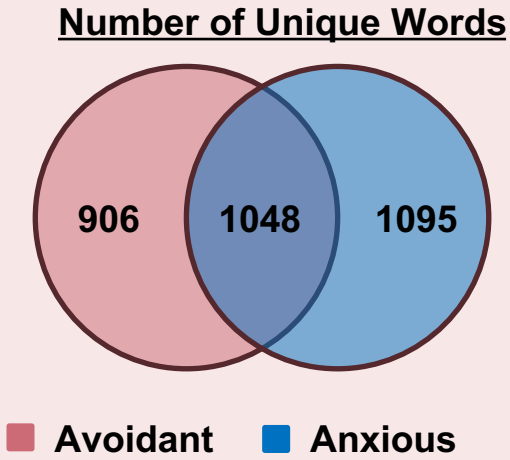
Text from each post are processed to sift out key features from each style.

Only text data (post title) are processed. Each post have an average of 186 words.

Each post will go through the following cleaning steps:

- Remove punctuations, URLs, numbers.
- Split the text into words (tokenize).
- Remove commonly used words (stop words) like “this”, “is”, “and”.
- Breaking down words into its root form (lemmatization) like “running” -> “run”.

About 50% of words found in one subreddit can also be found in another.



- For good NLP (Natural Language Processing) model performance, it is important for the text in both classes to be distinct.
- One challenge is that almost 50% of words found in one subreddit can be also found in another.

To improve model performance, overlapping words that appeared equally in both subreddits were removed.

Distribution of the Difference in Frequency of Overlapping Words

Percentile	Difference in Appearance Count	
50	1	Removed
60	2	Removed
70	3	Removed
80	5	Removed
90	8	Removed
92	10	Removed
94	12	Kept
96	20	Kept
98	29	Kept
100	699	Kept

- Not all overlapping words appeared equally in both subreddits. This difference in “appearance count” is a differentiating feature.
- Overlapping words that have a difference in appearance count of 10 or less were removed.

Examples of Overlapping Words

Overlapping Words	Appearance in "Avoidant"	Appearance in "Anxious"	Difference in Appearance Count	
legitimately	1	1	0	Removed
unattractive	1	1	0	Removed
emotional	15	10	5	Removed
deactivation	40	5	35	Kept
attachment	101	196	95	Kept

Model Overview

Data Class

- Class 0 – Avoidant Attachment Style (1,355 rows)
- Class 1 – Anxious Attachment Style (1,811 rows)

Vectorization Methods

- Count Vectorization
- N-Gram Vectorization
- TF-IDF (Term Frequency-Inverse Document Frequency)

Machine Learning Models Used

- Naive Bayes (Bernoulli)
- Naive Bayes (Multinomial)
- Logistic Regression

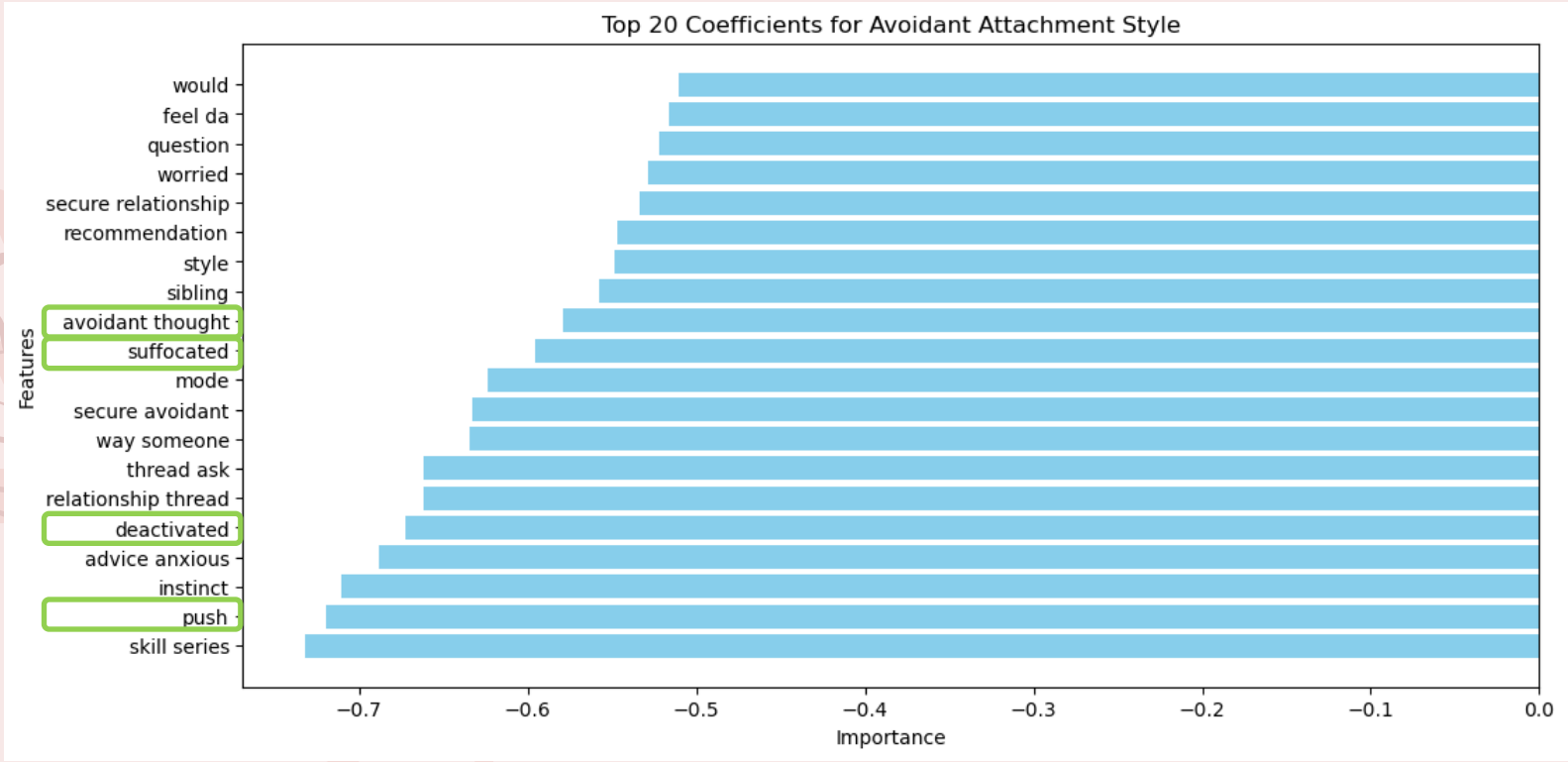
Vectorization is the process of encoding text as integers to create a feature vector of numerical features that represent an object.

Model Performance

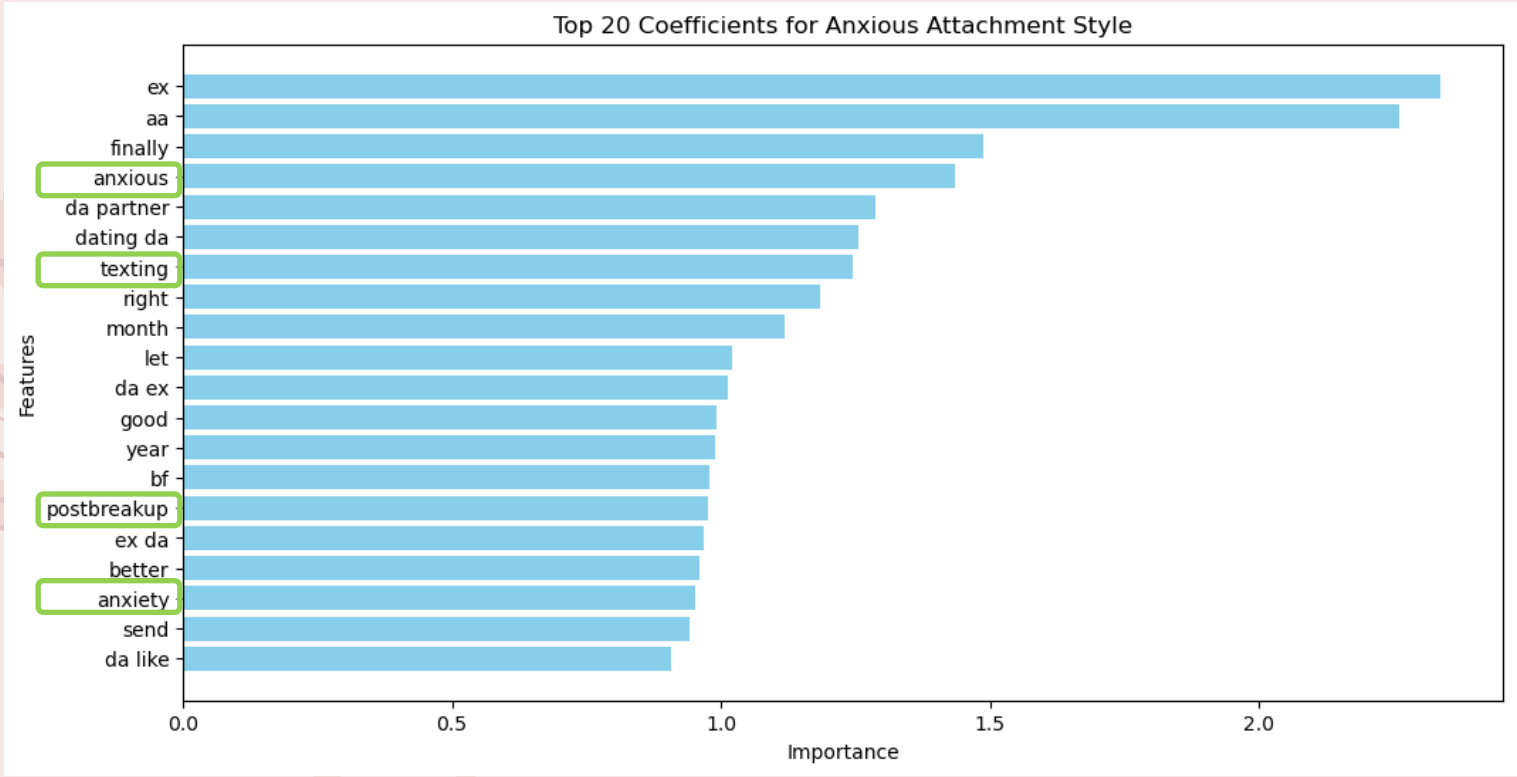
Models	Train accuracy	Test accuracy	Precision	Recall	F1 Score
Naive Bayes Bernoulli (With Count Vectorizer)	0.90	0.88	0.85	0.96	0.90
Naive Bayes Multinomial (With Count Vectorizer)	0.90	0.88	0.87	0.92	0.90
Logistic Regression (with Count Vectorizer)	0.90	0.90	0.89	0.93	0.91
Logistic Regression (with TF-IDF Vectorizer)	0.88	0.89	0.87	0.94	0.90

Accuracy = Total Correct Predictions / Total Predictions
Precision = Total Anxious Styles Predicted Correctly / Total Anxious Styles Predicted
Recall = Total Anxious Styles Predicted Correctly / All Anxious Styles Observed
F1 = Weighted Average of Precision and Recall

Top 20 features of Avoidant Style Class (based on coefficient value)



Top 20 features of Anxious Style Class (based on coefficient value)



Demo

attachment-style-classifier.streamlit.app

Summary

- **Problem statement:**

  How can we help couples facing difficulties in their relationship, strengthen communication, through recognising and understanding their attachment styles?

- **Deliverables:**
  - Build a classifier model with accuracy of > 0.8
  - Deploy an easy-to-use self-help tool using the classifier model embedded to provide information and resources for couples to improve on their communication with each other.

Moving forward

Phase 1

First release of the app on:

- MSF website
- MSF's Telegram channel (i.e.,MSFCares)

Phase 2

Improve accuracy of classifier model to ≥0.95 through:

- collecting more data from other similar subreddits
- exploring further ways to tune the existing classifier model
- exploring the use of other classifier models (e.g., XGBoost)

Phase 3

Second release of the app with:

- expanded coverage of the other two attachment styles

Thank you