Scan QR code to try the app!

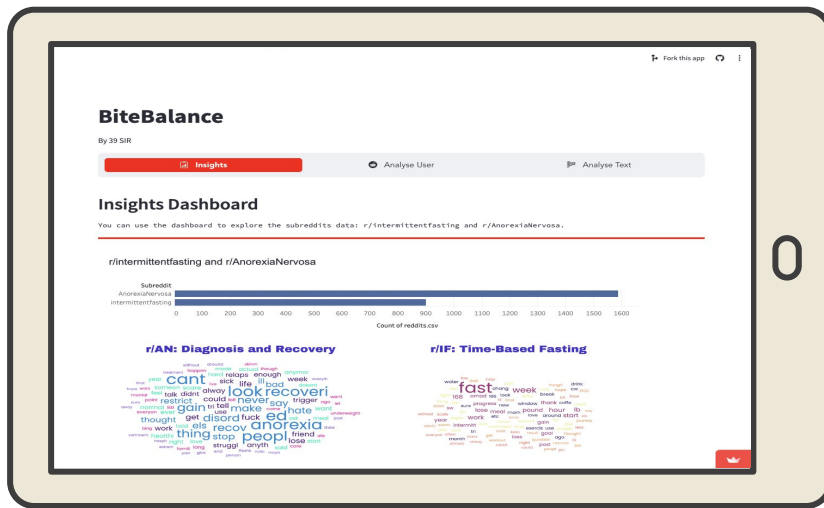# Table of Contents

Problem Statement

Dear **40 SIB Wellness** and **Corporate Partners**,
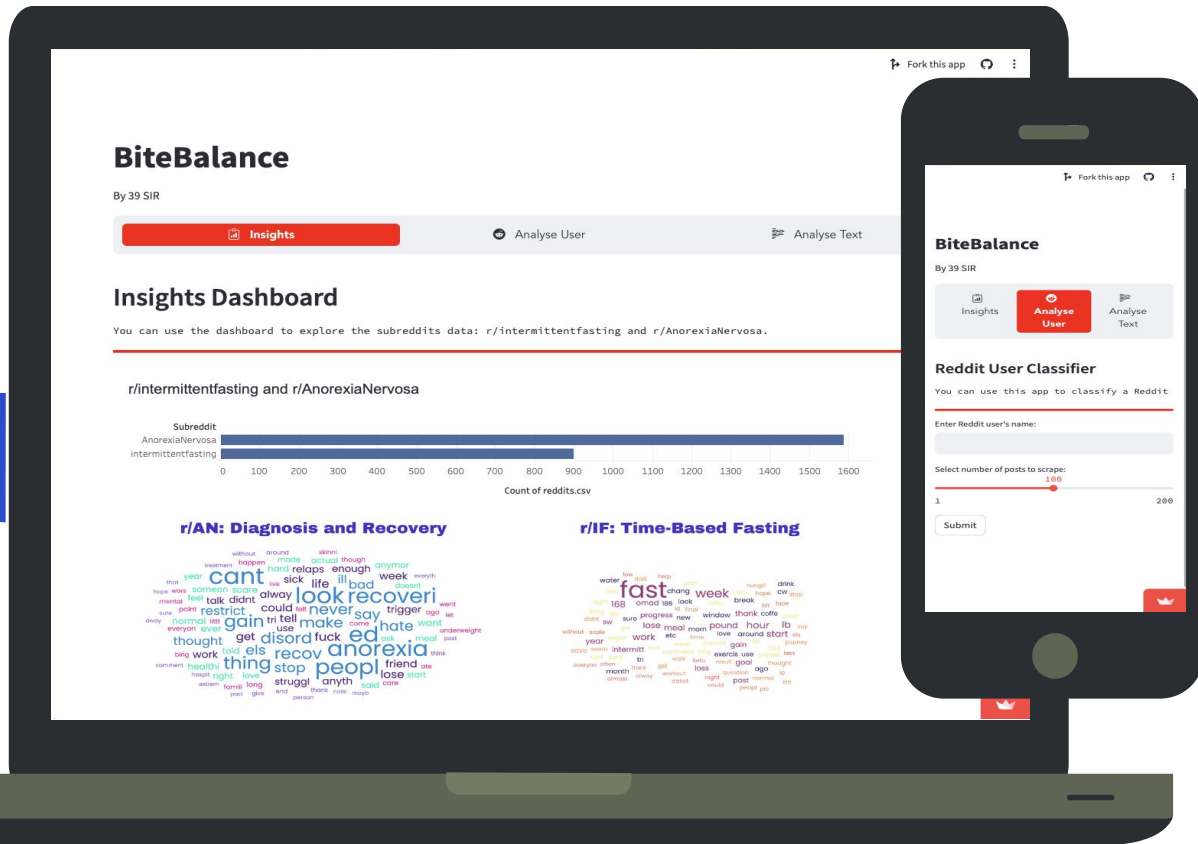
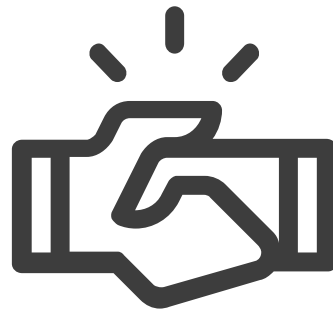Welcome to the launch of **BiteBalance** by **39 SIR Health**

BiteBalance for You

Text-based classifier that predicts between **intermittent fasting** and **anorexia nervosa**

For clients
Investigation of early signs of disordered eating

For partners
Lead generation and referrals for health and wellness support

Background

Intermittent Fasting (IF)

A popular health trend:

- Controlled fasting and eating

- For weight management and wellness

Anorexia Nervosa (AN)

A severe mental health disorder:

- Extreme dietary restriction

- Fear of gaining weight

- Distorted body image

Sources:
https://www.hopkinsmedicine.org/health/wellness-and-prevention/intermittent-fasting-what-is-it-and-how-does-it-work
https://www.aware.org.sg/information/eating-disorders

Extreme Intermittent Fasting?

12:12

16:8

OMAD:
One Meal A Day

12/12
Fasting

12 Hour
Fasting window

12 Hour
Eating window

8/16
Fasting

16 Hour
Fasting window

8 Hour
Eating window

OMAD
Fasting
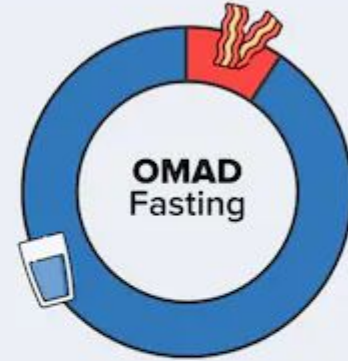
23 Hour
Fasting window

1 Hour
Eating window

Source: https://www.carbmanager.com/article/yoherxeaaceazayu/how-to-choose-an-intermittent-fasting-schedule

Key Question

How can we **differentiate** between users from **r/intermittentfasting** and **r/AnorexiaNervosa** based on their Reddit posts?

Our PRAW-rocess

Web Scraping

01 → **02**

Data Cleaning & EDA

3960 posts from **r/intermittentfasting** and **r/AnorexiaNervosa**

Removal of duplicates, and combining of title and text

Model and Evaluation

03 → **04**

App Development

3 classifiers and 3 vectorizers

Reddit Habits: Text in Title Only
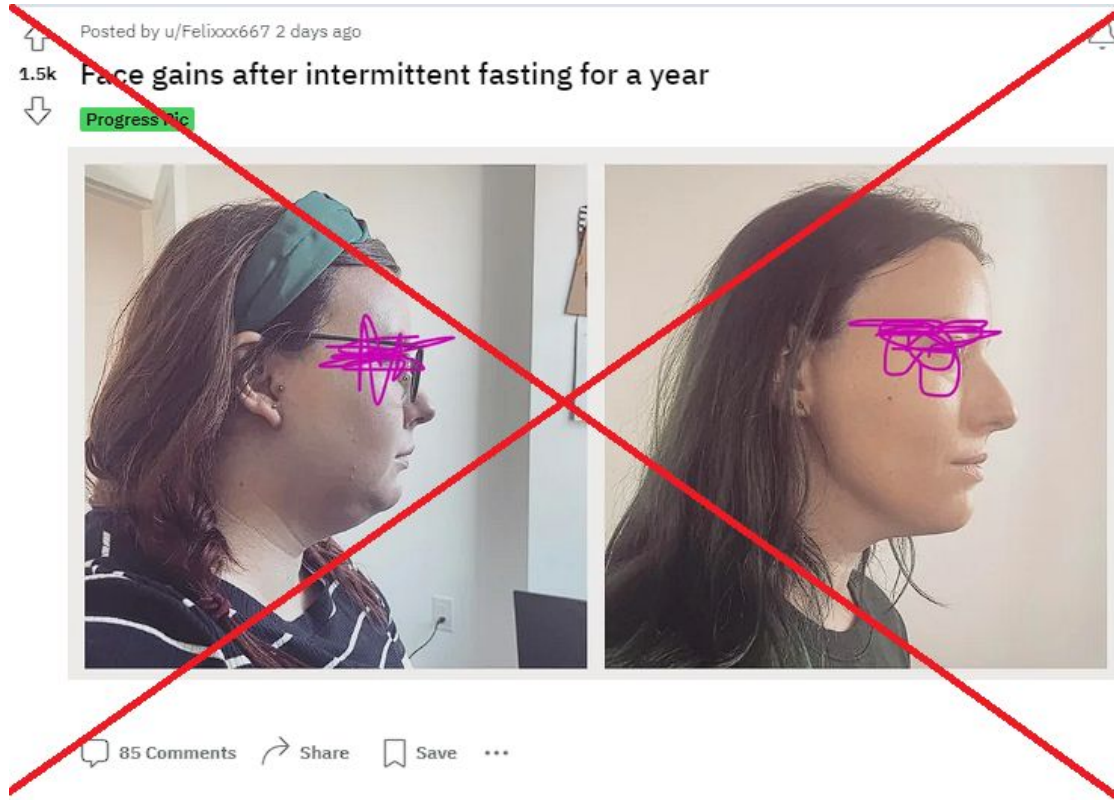
Reddit Habits: Text in Title and Body

11 **How do you, personally, get yourself back into it after stopping for a while?**

`Seeking Advice`

I got into my goal range (f, 5'7", 150, goal weight 145) with IF last Fall, and then entered into a very miserable period in which two of our kids had started school and were annihilating us with rapid fire illnesses. We were continuously sick for over three months and require multiple antibiotics and inhalers. During this time, I stopped fasting because eating gave me a much-needed energy boost. Then, when I was finally better and ready to get back at it (three pounds heavier btw), I just kind of couldn't. It used to be so easy, but now it's hard to consistently skip breakfast (even harder to skip dinner - skipping lunch seems like an option to possibly consider). I'm a super active person and usually walk like five miles by lunch time, and usually more after that. Lately I sometimes feel faint and start sweating bullets if I skip breakfast. Btw, for context, none of the illnesses were covid, so it's probably not long-covid-related.

Recently I got back into a pretty good groove for a few weeks, but then I fell off the wagon due to cycle-related stuff (you know... Menstrual munchies, mostly caused by an increase in that faint feeling). Now I'm having a hard time again.

What do you guys do to get back on track? Obviously self-discipline would be great, but it feels like my body chemistry is resisting IF and reacting poorly. Maybe there are specific (more specific than "self-discipline") tricks for getting everything working again.

Followup query: who here has kids and can't seem to stop picking at the food you prepare for them that you know they're not gonna eat, anyway? I swear I'm putting away at least 300 extra calories a day nibbling on the damn muffins I only have around for the kids. Who's gotten past this issue?

Data Cleaning

DataFrame

- Removed Irrelevant Posts
- Deduplicated Data
- Standardized Column Headers
- Handled Missing Data
- Merged 'Title' and 'Post' Columns

Text

- Removed Punctuation
- Removed Stopwords
- Lowercase
- Drop Hyperlinks
- Stemming
- Lemmatization

Data Cleaning: Stemming produces greater variety than lemmatizing

IF (S)	IF (L)	AN (S)	AN(L)
fast	fasting	cant	cant
week	fast	look	ed
hour	week	ed	anorexia
work	hour	anorexia	people
lb	lb	peopl	recovery
start	168	recoveri	thing
168	pound	thing	never
lose	year	gain	look
pound	meal	recov	else
year	work	disord	hate

EDA: Longer text posts in r/AN compared to r/IF

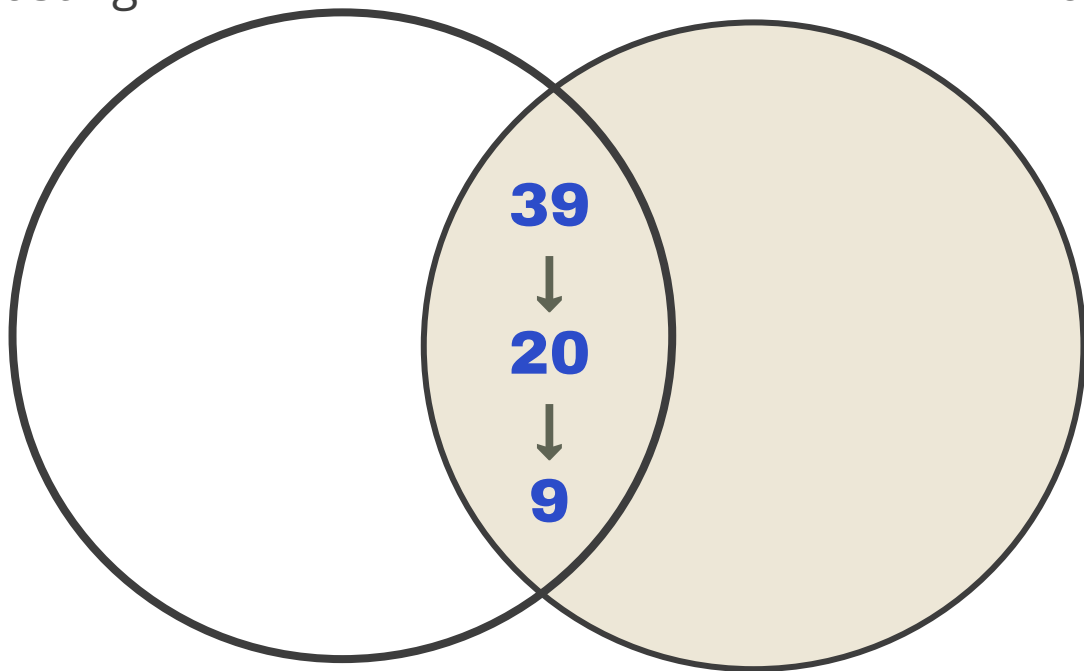**Distribution of Text Length (Outliers Removed vs Upper Outliers)**

r/IF: Time-Based Fasting

r/AN: Diagnosis and Recovery
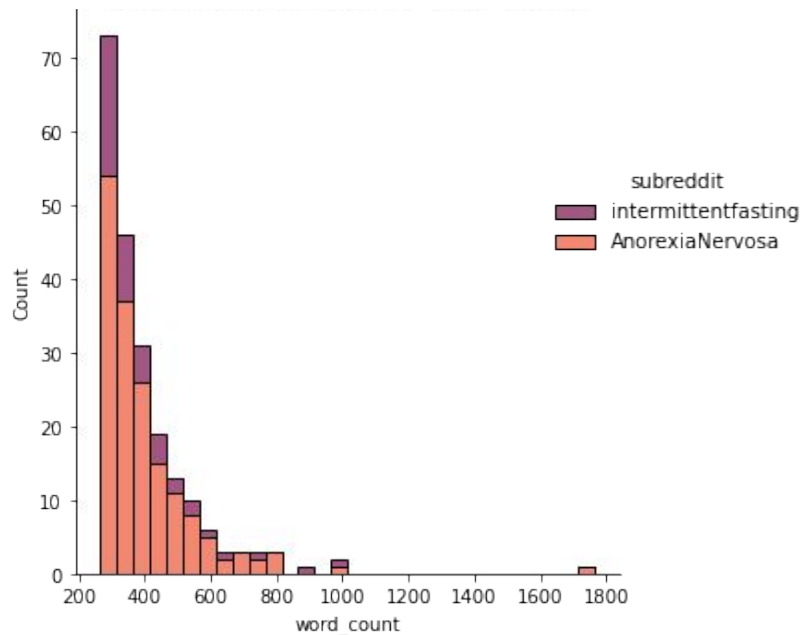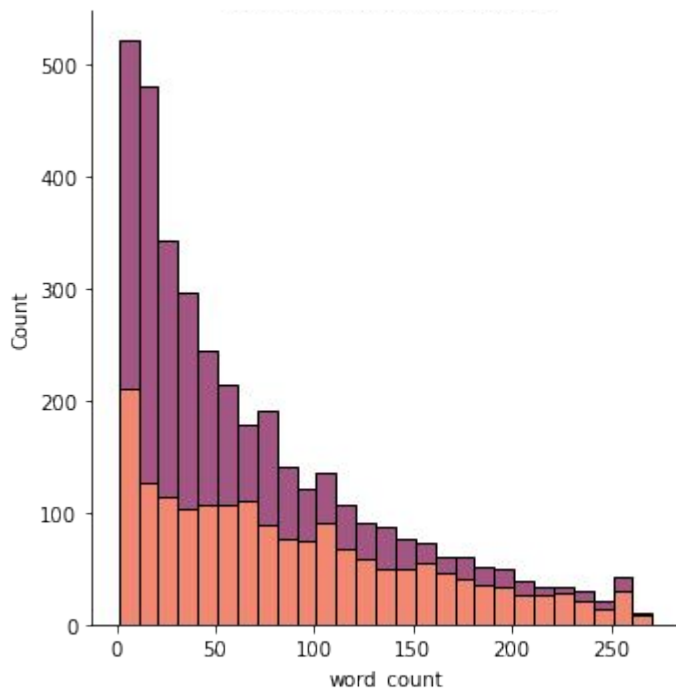
Creation of Stopwords

Top 100 Words
r/intermittentfasting

Top 100 Words
r/AnorexiaNervosa

39

↓

20

↓

9

r/IF: Time-Based Fasting

Intermittent Fasting	Anorexia
intermitt fast	sick enough
break fast	mental health
hour fast	ice cream
fast hour	anorexia nervosa
black coffe	extrem hunger
fast window	best friend
fast water	hate hate
type fast	high school
daili fast	diagnos anorexia
hour etc	struggl anorexia

r/AN: Discussion of Mental Health

Intermittent Fasting	Anorexia
intermitt fast	sick enough
break fast	mental health
hour fast	ice cream
fast hour	anorexia nervosa
black coffe	extrem hunger
fast window	best friend
fast water	hate hate
type fast	high school
daili fast	diagnos anorexia
hour etc	struggl anorexia

r/IF, r/AN: Food Keywords

Intermittent Fasting	Anorexia
intermitt fast	sick enough
break fast	mental health
hour fast	ice cream
fast hour	anorexia nervosa
black coffe	extrem hunger
fast window	best friend
fast water	hate hate
type fast	high school
daili fast	diagnos anorexia
hour etc	struggl anorexia

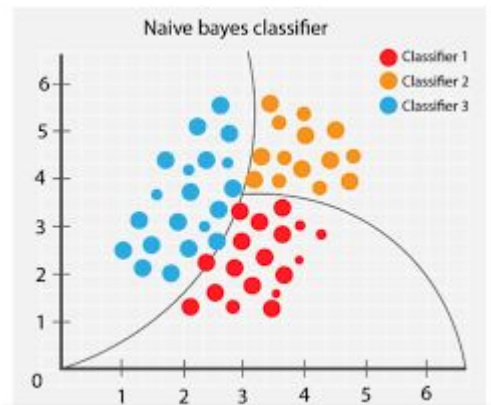# Classification Models

Logistic Regression
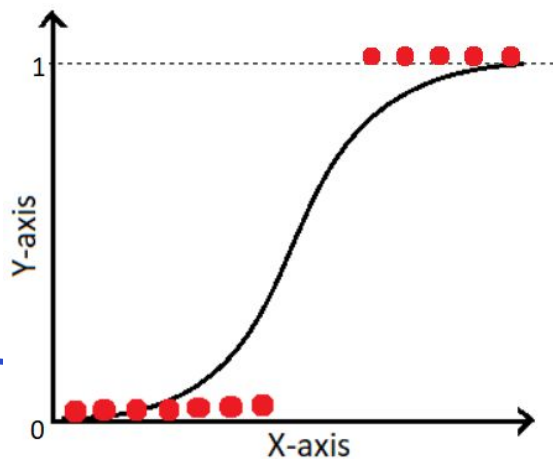
Graph

Naive Bayes


Boundaries

XGBoost
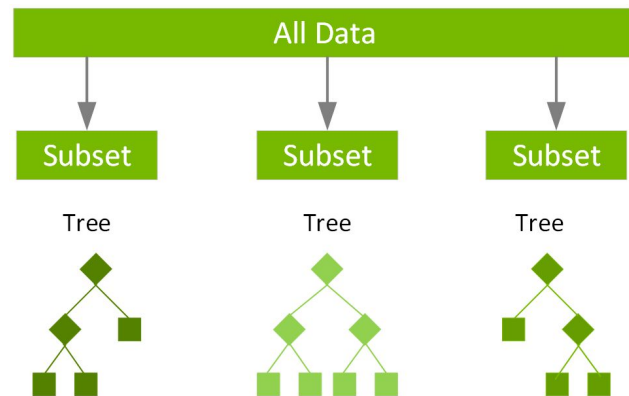
Decision Trees

Getting the best score

1. Feature Engineering

2. Compare Against Other Models

Baseline Model: Logistic Regression

Model	Vectorizer	Accuracy (train)	Accuracy (test)	Precision (test)	Recall (test)	F1 (test)	ROC (test)	Execution time (ms)
Logistic Regression	Count Vectorizer	0.96	0.92	0.94	0.9	0.92	0.92	10.33

Best Model Score for Each Round of Cleaning
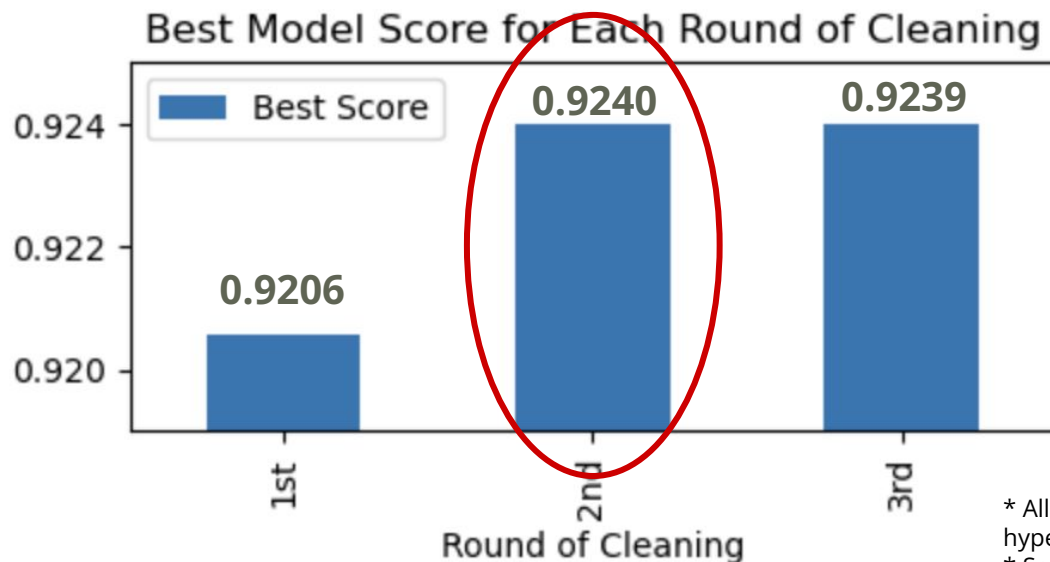
* All models have been optimised by tuning their hyperparameters before being run.
* Scores closer to 1 the better

Total of 9 Models Deployed

Logistic Regression	Multinomial Naive Bayes	XGBoost
CountVectorizer	CountVectorizer	CountVectorizer
Logistic Regression	**Multinomial Naive Bayes**	**XGBoost**
N-gram (2,2)	N-gram (2,2)	N-gram (2,2)
Logistic Regression	**Multinomial Naive Bayes**	**XGBoost**
TF-IDF	TF-IDF	TF-IDF

Deployed: Multinomial CV

Model	Vectorizer	Accuracy (train)	Accuracy (test)	Precision (test)	Recall (test)	F1 (test)	ROC (test)	Execution time (ms)
Logistic Regression	Count Vectorizer	0.96	0.92	0.94	0.9	0.92	0.92	10.33
XGBoost	N-gram (2,2)	0.85	0.79	0.88	0.64	0.74	0.78	810.11
Logistic Regression	N-gram (2,2)	0.94	0.81	0.93	0.66	0.77	0.80	49.11
Multinomial NB	N-gram (2,2)	0.98	0.90	0.90	0.89	0.89	0.9	1.88
Logistic Regression	TF-IDF	0.99	0.93	0.93	0.91	0.92	0.93	9.71
Multinomial NB	TF-IDF	0.96	0.93	0.91	0.94	0.93	0.93	1.03
XGBoost	TF-IDF	0.98	0.93	0.91	0.94	0.93	0.93	794.21
XGBoost	Count Vectorizer	0.97	0.94	0.92	0.95	0.94	0.94	165.05
Multinomial NB	Count Vectorizer	0.96	0.95	0.95	0.94	0.94	0.95	0.96

App Demo

Conclusion

Deployment of **Multinomial NB** model with **0.95** accuracy

r/IF Posts

- **Shorter** text posts
- Content on **time-based** fasting
- Food keywords - e.g. **black coffee**

r/AN Posts

- **Longer** text posts
- Discussion of **mental health** concerns
- Food keywords - e.g. **ice cream**

BiteBalance Recommender: Products and Services

Intermittent Fasting

- **Gym Equipment** by Colony.co

- **Protein Supplements** by Protein la Muffins

- **Grooming Classes** by 79 @ Anson

Anorexia Nervosa

- **Stress Management** by 40 SIB Wellness

Next Steps

Share

Share with your clients the differences between intermittent fasting and symptoms of anorexia

Investigate

Keep an eye out for clients practicing intermittent fasting concerned about mental health

Review

39 SIR will share initial findings after 30 days to determine market size

BiteBalance v2

Cloud-Based Data Storage

Protection of client data as user-base grows

Social Media Integration

Scrape platforms that have lesser moderation than Reddit

Image Recognition

Training of the image-only posts from Reddit on our model

Thank you.