Kyera Francis
Data Scientist

Analyzing Community Sentiments in Online Tribes: A Natural Language Processing (NLP) Approach

Military and Mental Health Analysis

Speaker:
Kyera Francis

Table Of Content

Primary Goal:
To classify communities using NLP techniques and to identify and analyze sentiments within specific communities.

01 Introduction

02 Military Classification

03 Military Sentiment

04 Mental Health Classification

05 Mental Health Sentiment

06 Trend of Techonology

07 Future Application

08 Conclusions & Rec'mds

09 Q&A

Introduction

Baracks Banter: Decoding the Army-Marine Lexicon.

- SGT Francis, Kyera

Achievement:

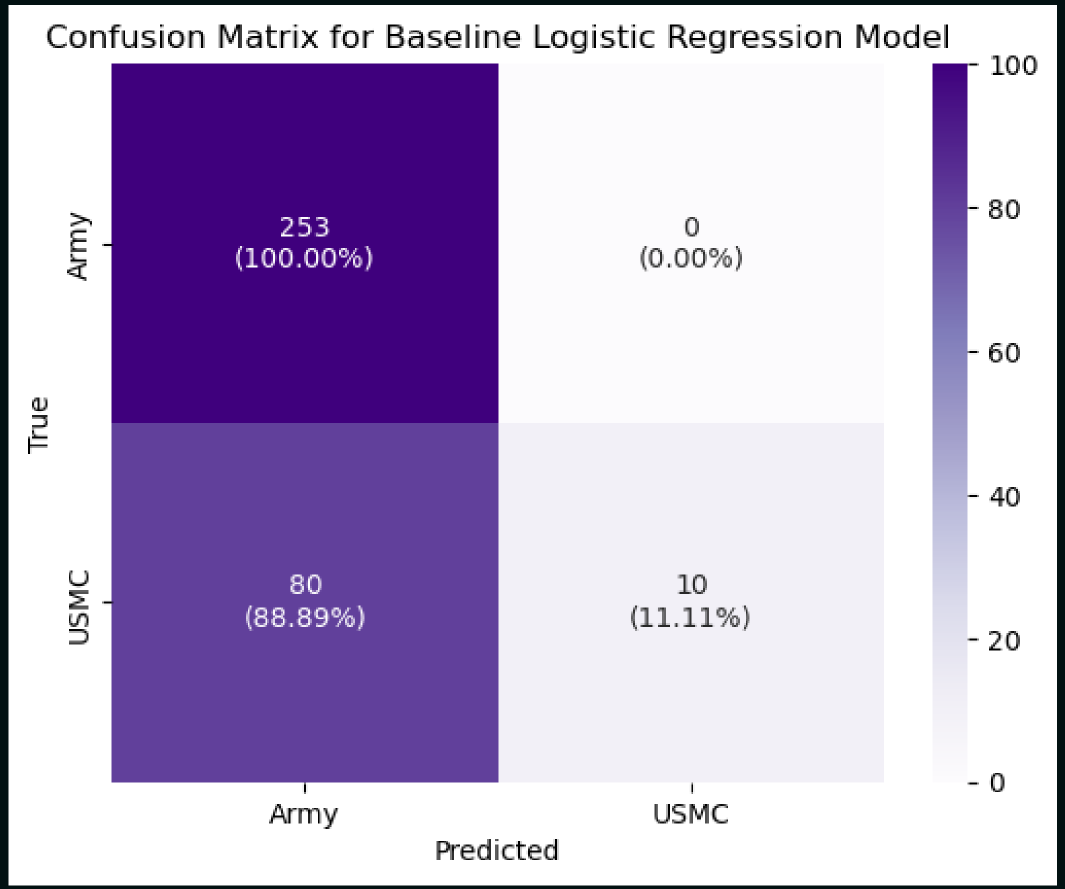
NLP Military Analysis

US Army vs US Marine Corps

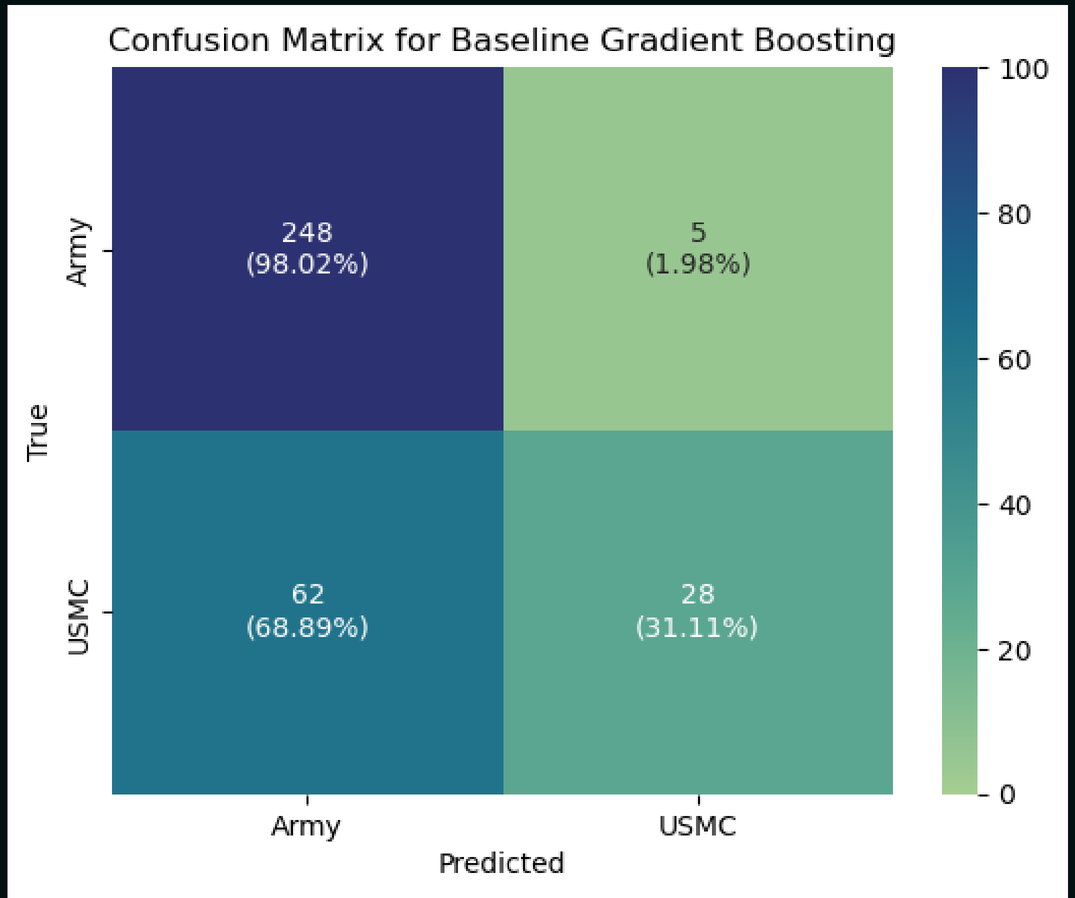
NLP Models: TF - IDF +
Logistical Regression and Gradient Boosting

Sentiment Model: TextBlob

BASELINE



	precision	recall	f1-score	support
Army	0.76	1.00	0.86	253
USMC	1.00	0.11	0.20	90
accuracy			0.77	343
macro avg	0.88	0.56	0.53	343
weighted avg	0.82	0.77	0.69	343



	precision	recall	f1-score	support
Army	0.80	0.98	0.88	253
USMC	0.85	0.31	0.46	90
accuracy			0.80	343
macro avg	0.82	0.65	0.67	343
weighted avg	0.81	0.80	0.77	343

NLP Military Analysis

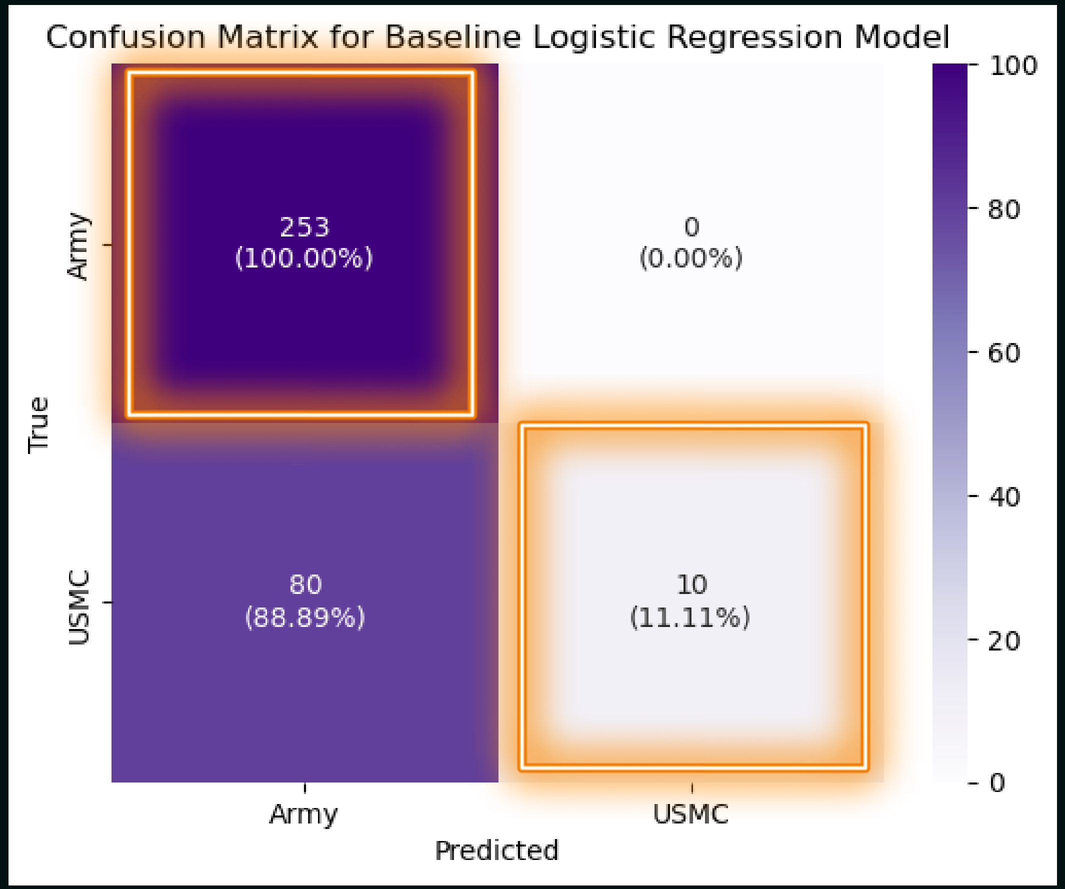
US Army vs US Marine Corps

NLP Models: TF - IDF +

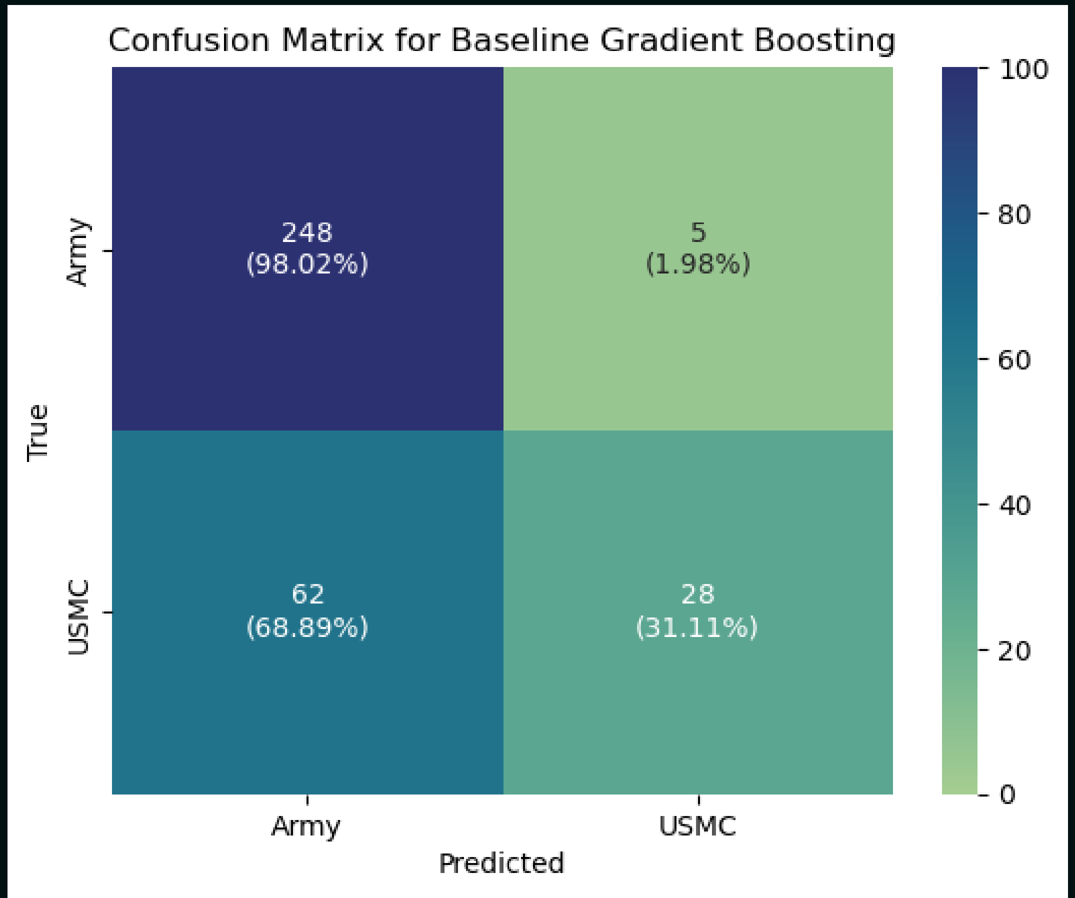
Logistical Regression and Gradient Boosting

Sentiment Model: TextBlob

BASELINE



	precision	recall	f1-score	support
Army	0.76	1.00	0.86	253
USMC	1.00	0.11	0.20	90
accuracy			0.77	343
macro avg	0.88	0.56	0.53	343
weighted avg	0.82	0.77	0.69	343



	precision	recall	f1-score	support
Army	0.80	0.98	0.88	253
USMC	0.85	0.31	0.46	90
accuracy			0.80	343
macro avg	0.82	0.65	0.67	343
weighted avg	0.81	0.80	0.77	343

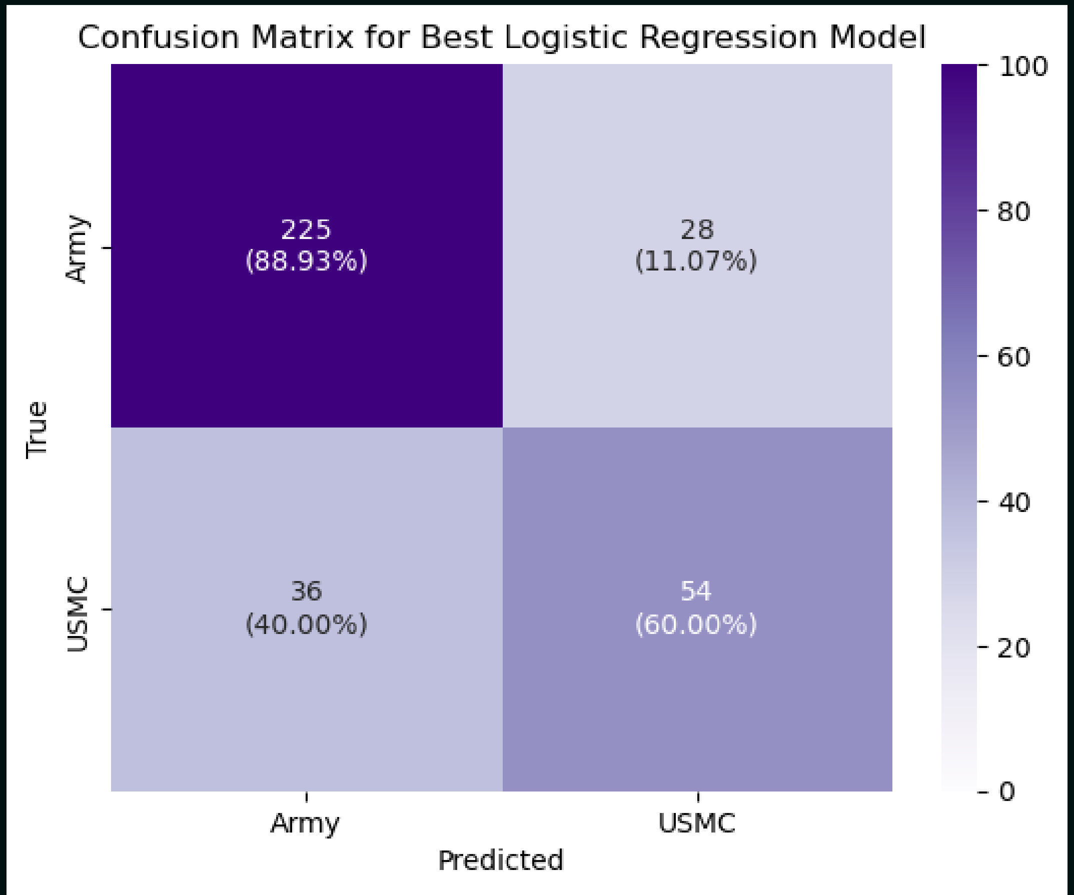
NLP Military Analysis

US Army vs US Marine Corps

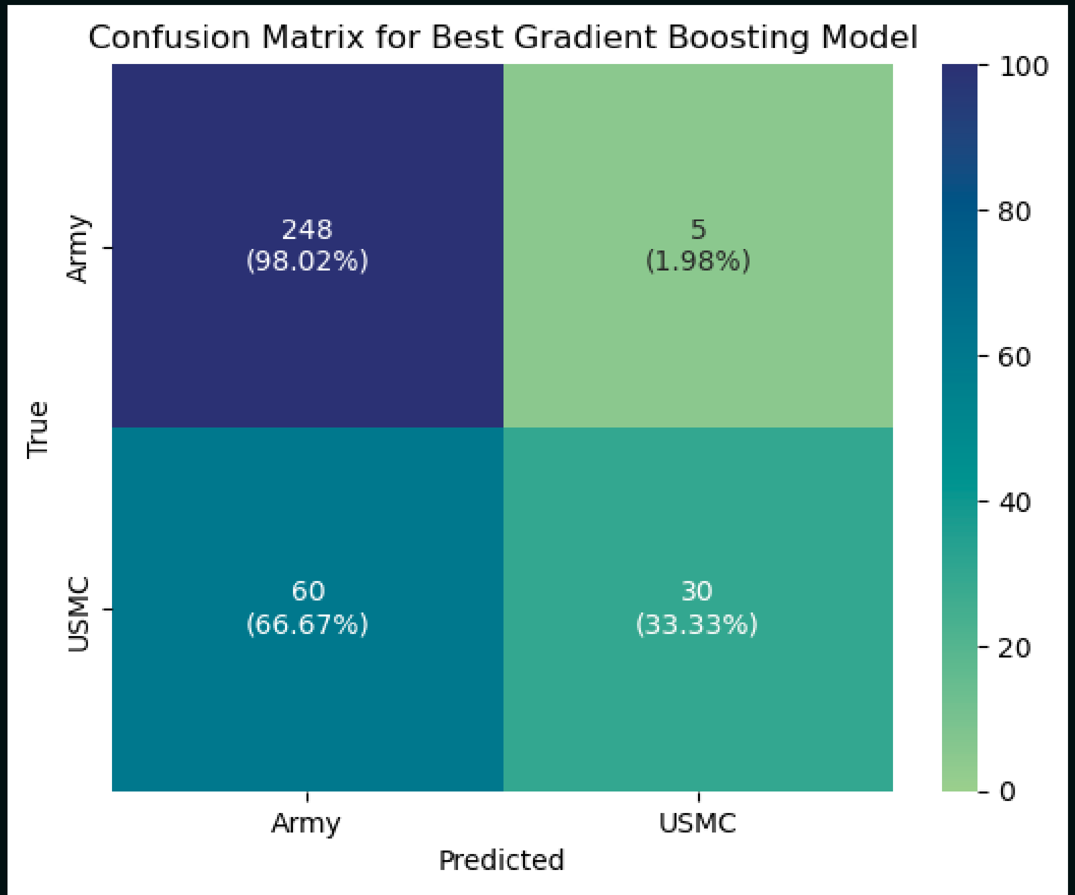
NLP Models: TF - IDF +
Logistical Regression and Gradient Boosting

Sentiment Model: TextBlob

BEST MODEL



	precision	recall	f1-score	support
Army	0.86	0.89	0.88	253
USMC	0.66	0.60	0.63	90
accuracy			0.81	343
macro avg	0.76	0.74	0.75	343
weighted avg	0.81	0.81	0.81	343



	precision	recall	f1-score	support
Army	0.81	0.98	0.88	253
USMC	0.86	0.33	0.48	90
accuracy			0.81	343
macro avg	0.83	0.66	0.68	343
weighted avg	0.82	0.81	0.78	343

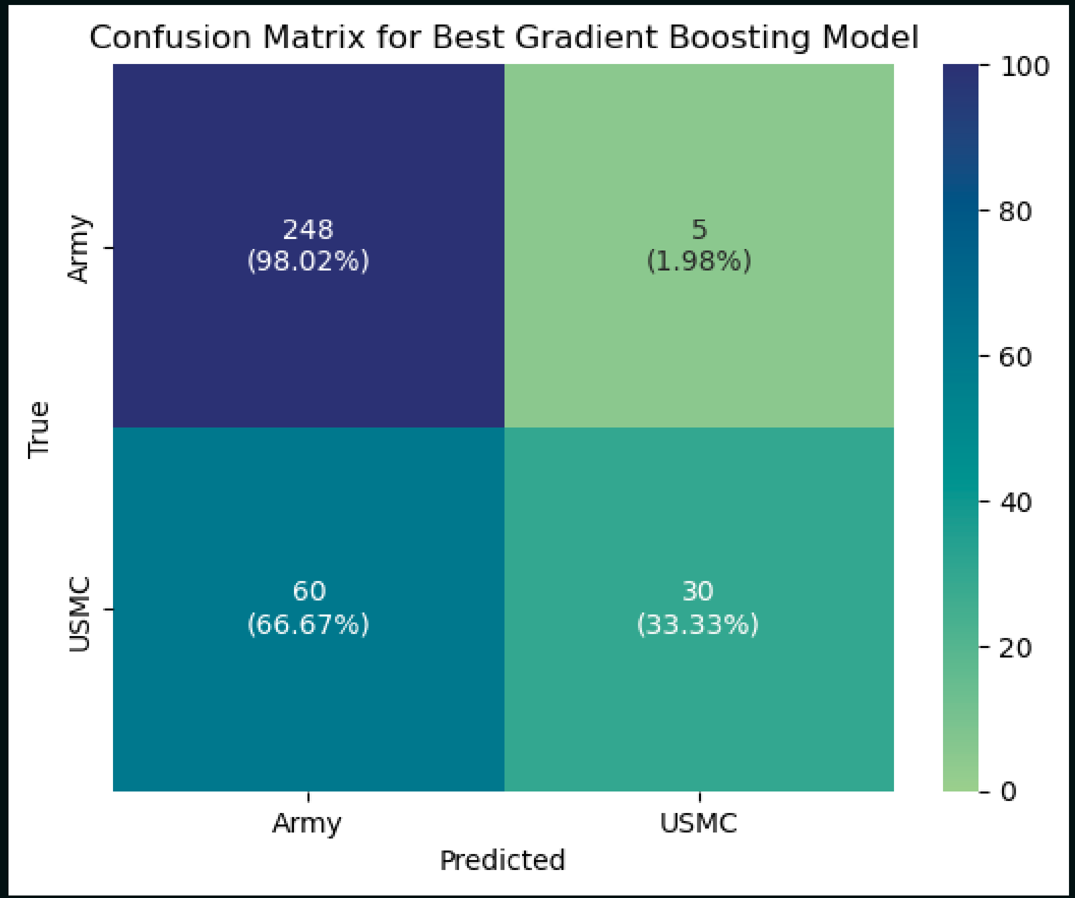
NLP Military Analysis

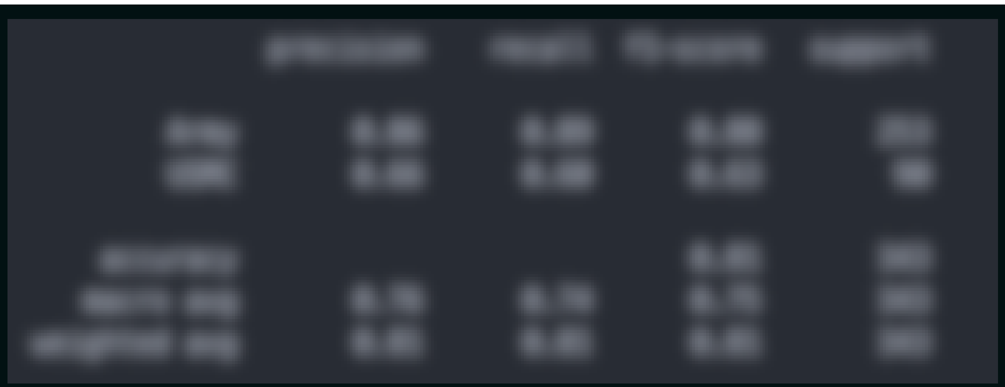
US Army vs US Marine Corps

NLP Model: TF - IDF + Gradient Boosting

Sentiment Model: TextBlob

BEST MODEL





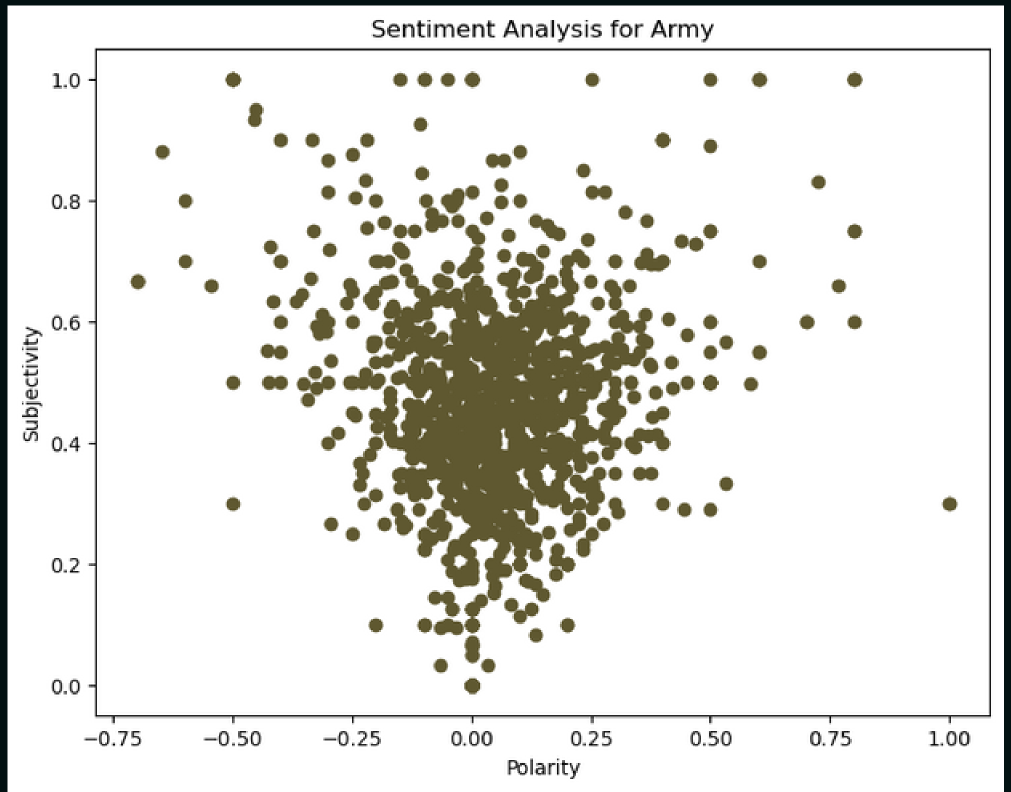
	precision	recall	f1-score	support
Army	0.81	0.98	0.88	253
USMC	0.86	0.33	0.48	90
accuracy			0.81	343
macro avg	0.83	0.66	0.68	343
weighted avg	0.82	0.81	0.78	343

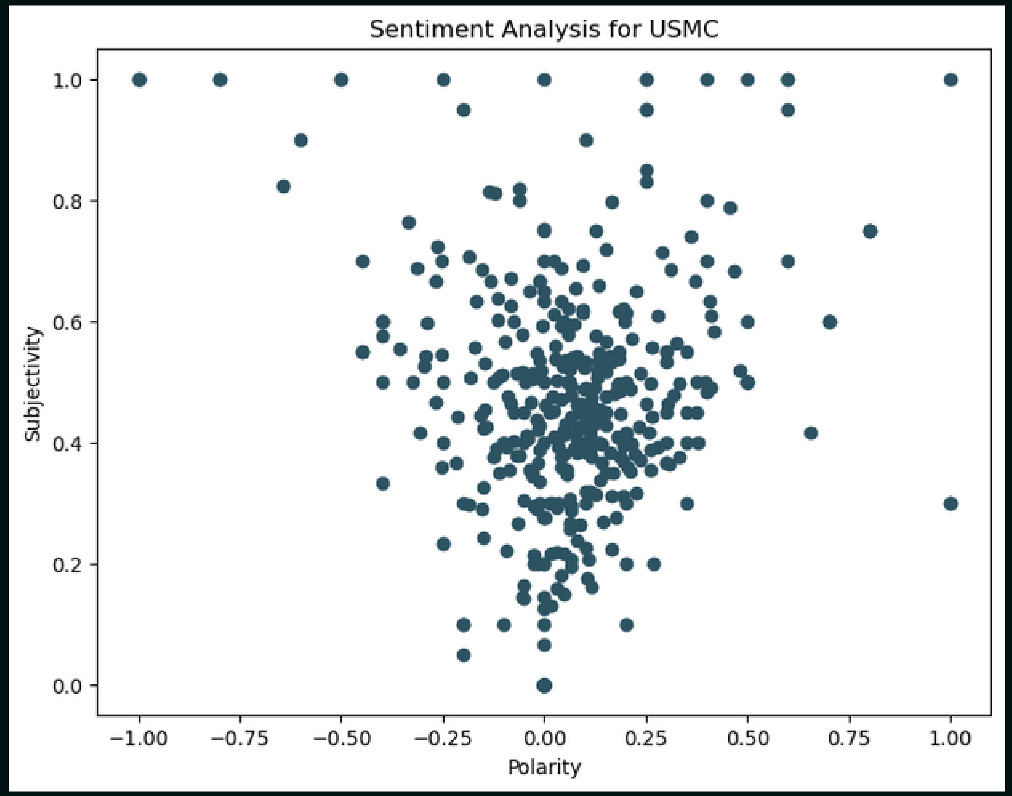
NLP Military Analysis

US Army vs US Marine Corps

NLP Model: TF- IDF + Gradient Boosting

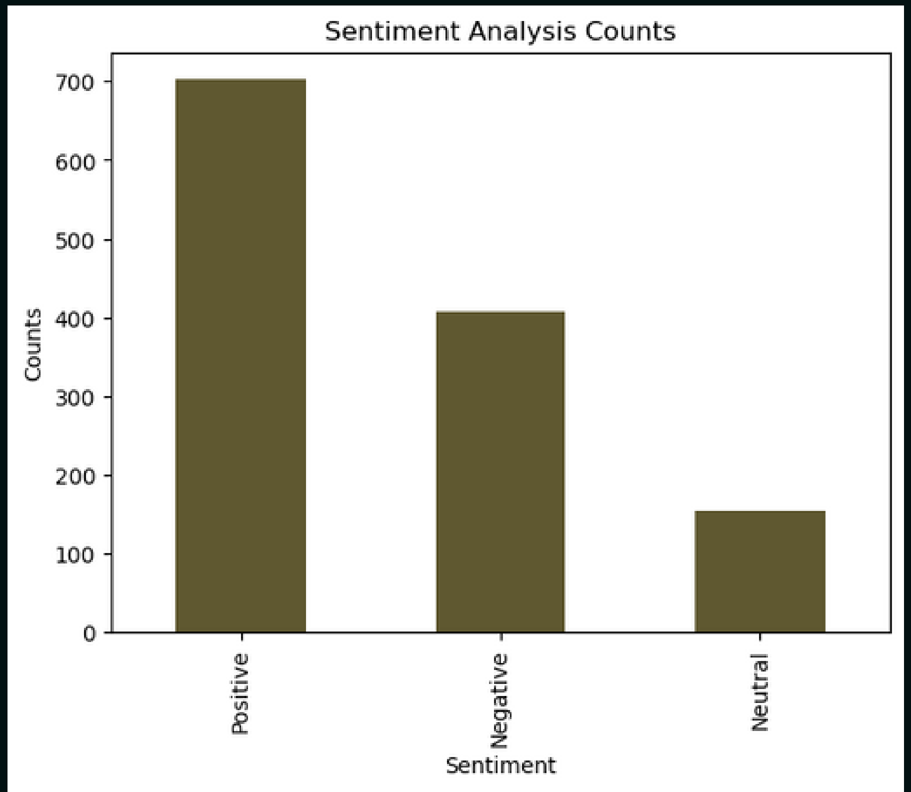
Sentiment Model: TextBlob





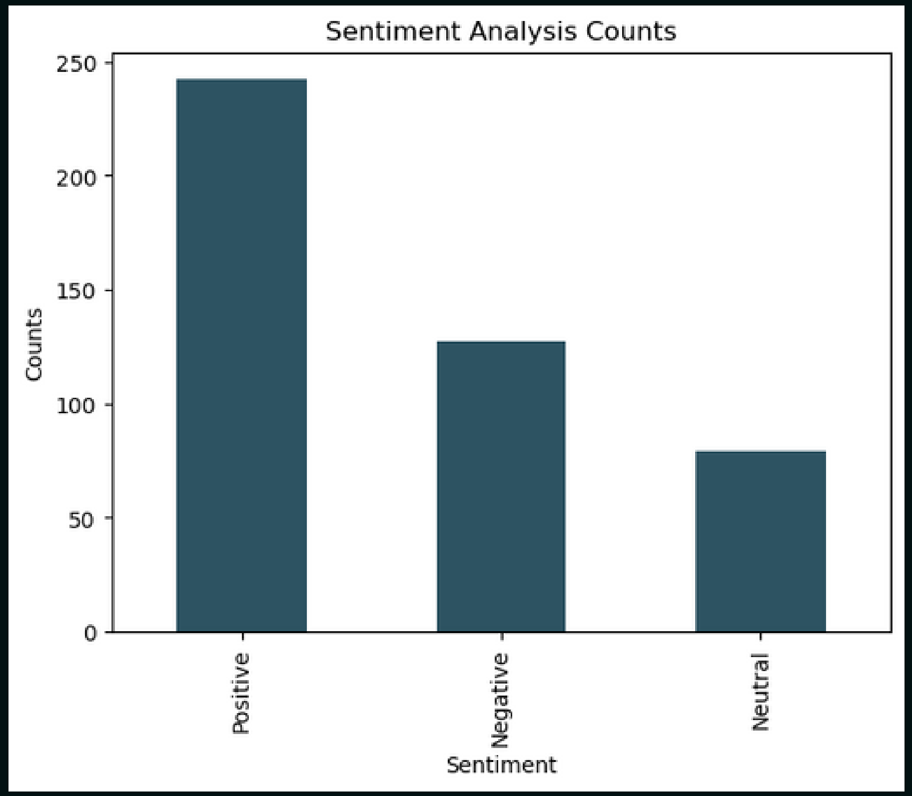
Army

```
Positive Posts Percentage: 55.5%
Negative Posts Percentage: 32.2%
Neutral Posts Percentage: 12.3%
```



Marine

```
Positive Posts Percentage: 54.0%
Negative Posts Percentage: 28.3%
Neutral Posts Percentage: 17.6%
```



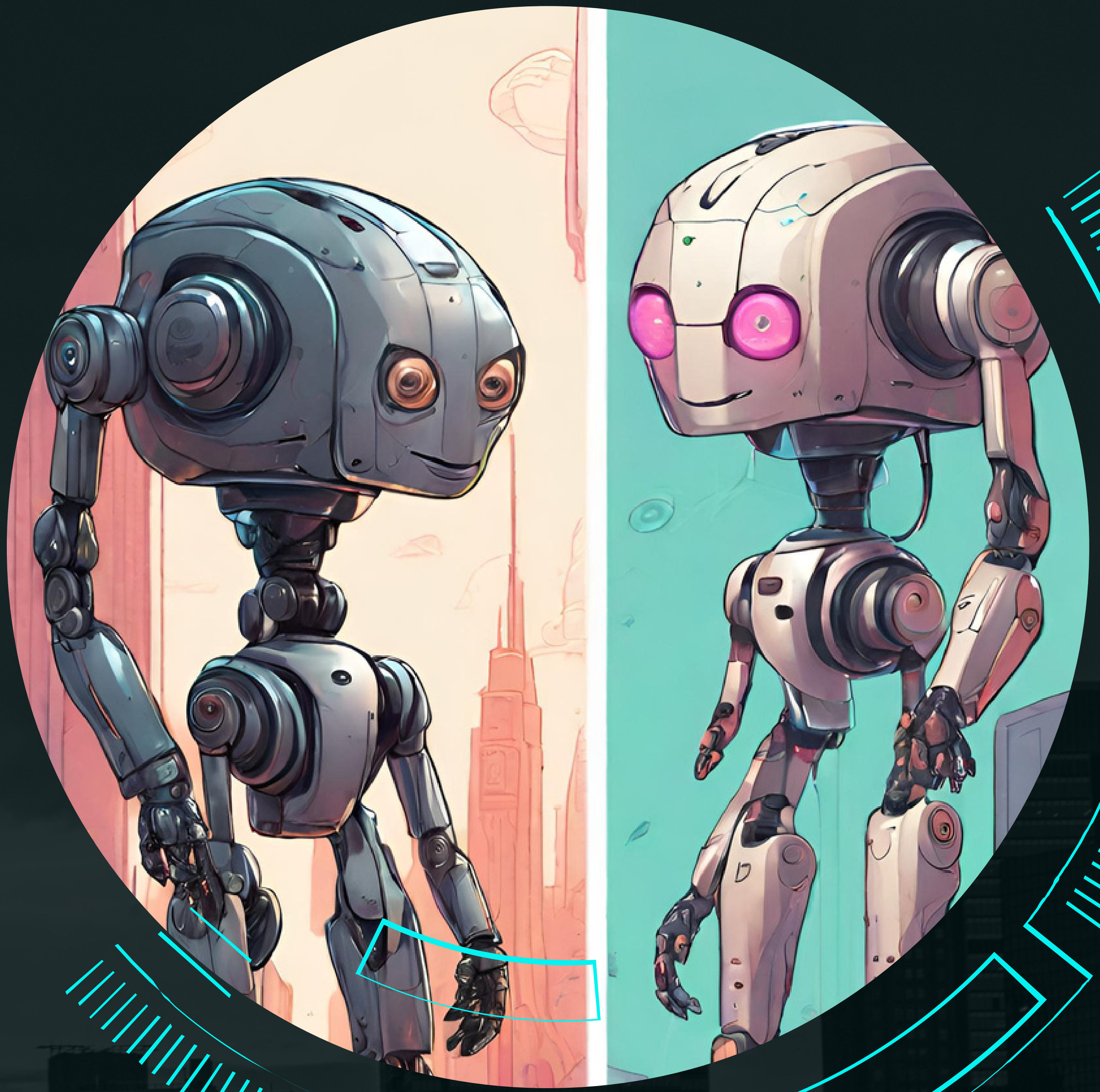
Military vs. Mental Health

How efficient are our models?

Can it perform effectively on another similiar subgroup of a community?

TF-IDF + Gradient Boosting Classifier

TextBlob



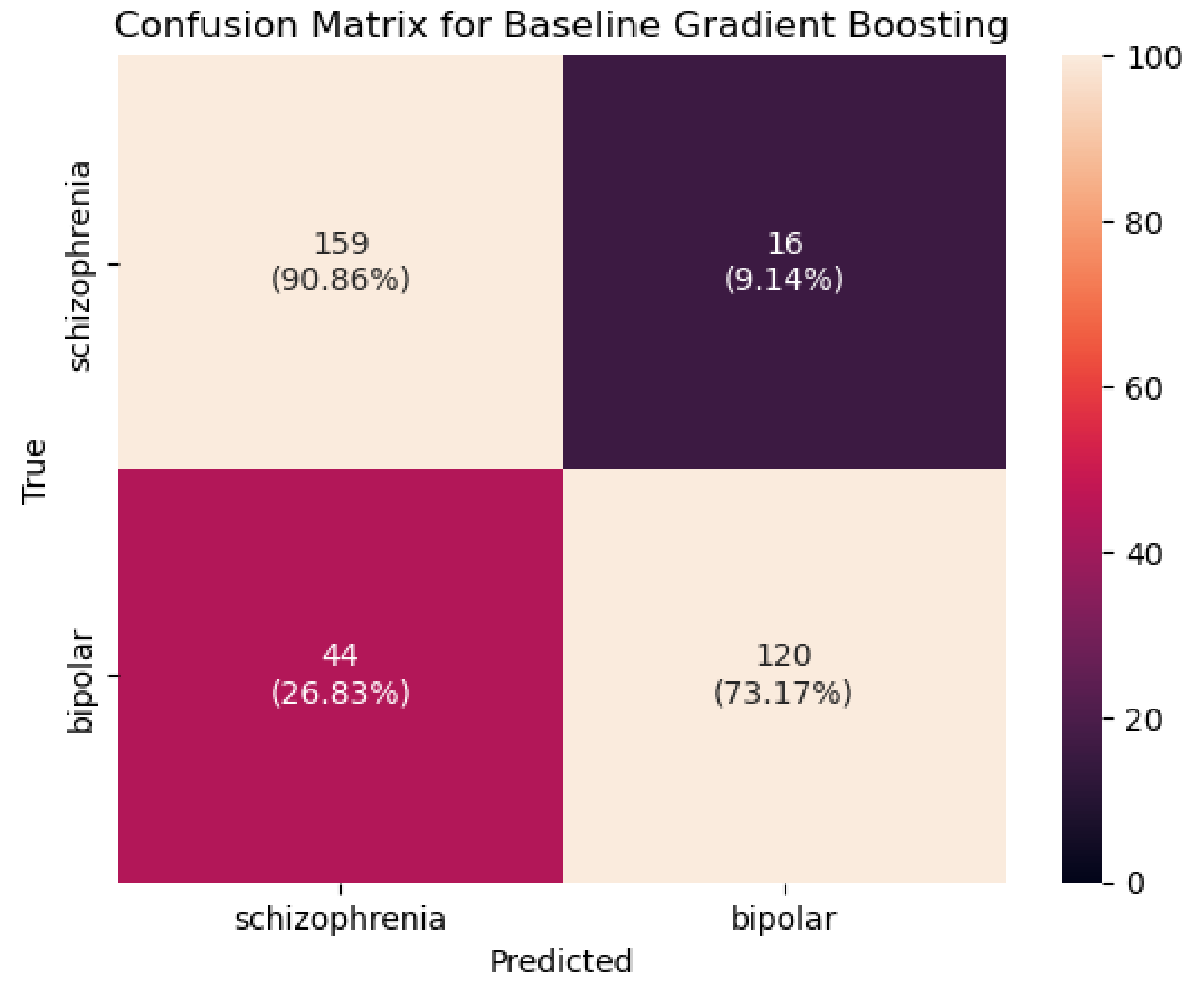
NLP M. Health Analysis

Schizophrenic vs Bipolar

NLP Model: TF - IDF + **Gradient Boosting**

Sentiment Model: TextBlob

BASELINE



	precision	recall	f1-score	support
Class				
bipolar	0.78	0.91	0.84	175.0
schizophrenia	0.88	0.73	0.80	164.0
accuracy			0.82	
macro avg	0.83	0.82	0.82	339.0
weighted avg	0.83	0.82	0.82	339.0

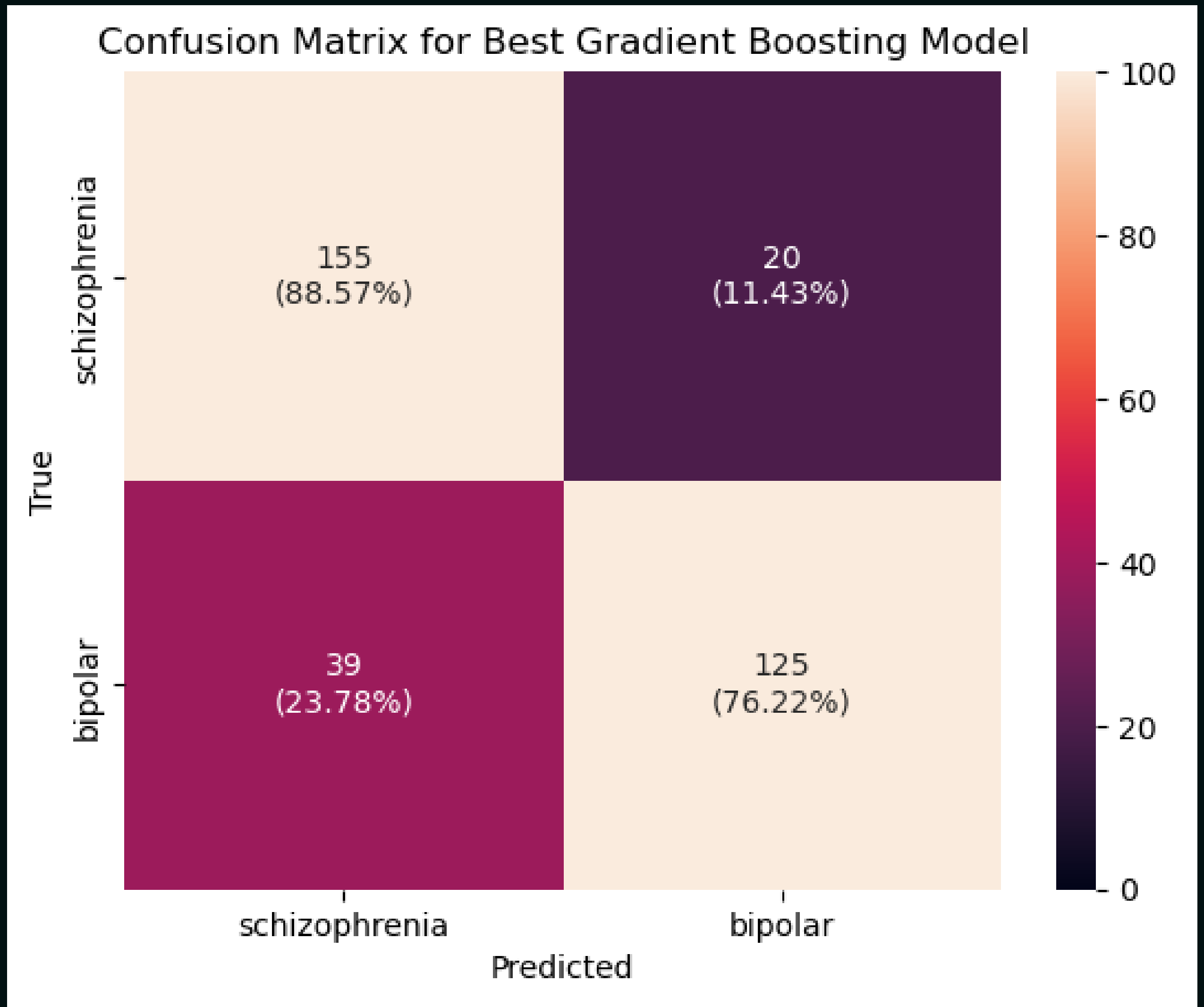
NLP M. Health Analysis

Schizophrenic vs Bipolar

NLP Model: TF - IDF + **Gradient Boosting**

Sentiment Model: TextBlob

BEST MODEL



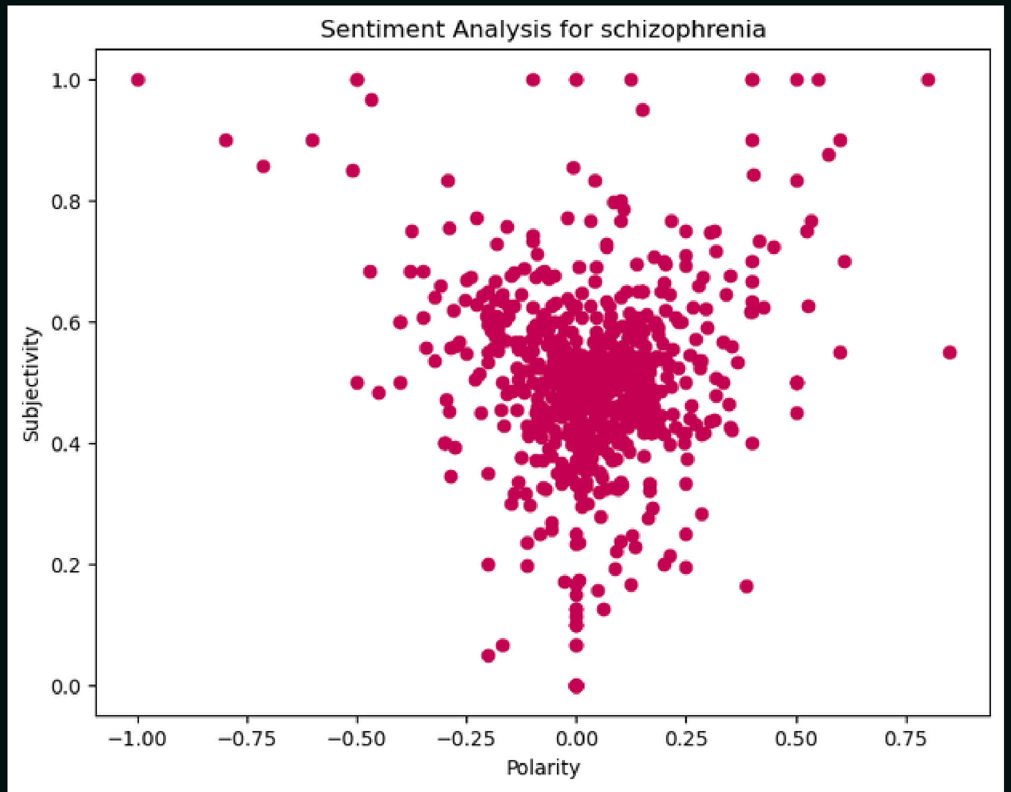
Class	precision	recall	f1-score	support
bipolar	0.8	0.89	0.84	175.0
schizophrenia	0.86	0.76	0.81	164.0
accuracy			0.83	
macro avg	0.83	0.82	0.82	339.0
weighted avg	0.83	0.83	0.83	339.0

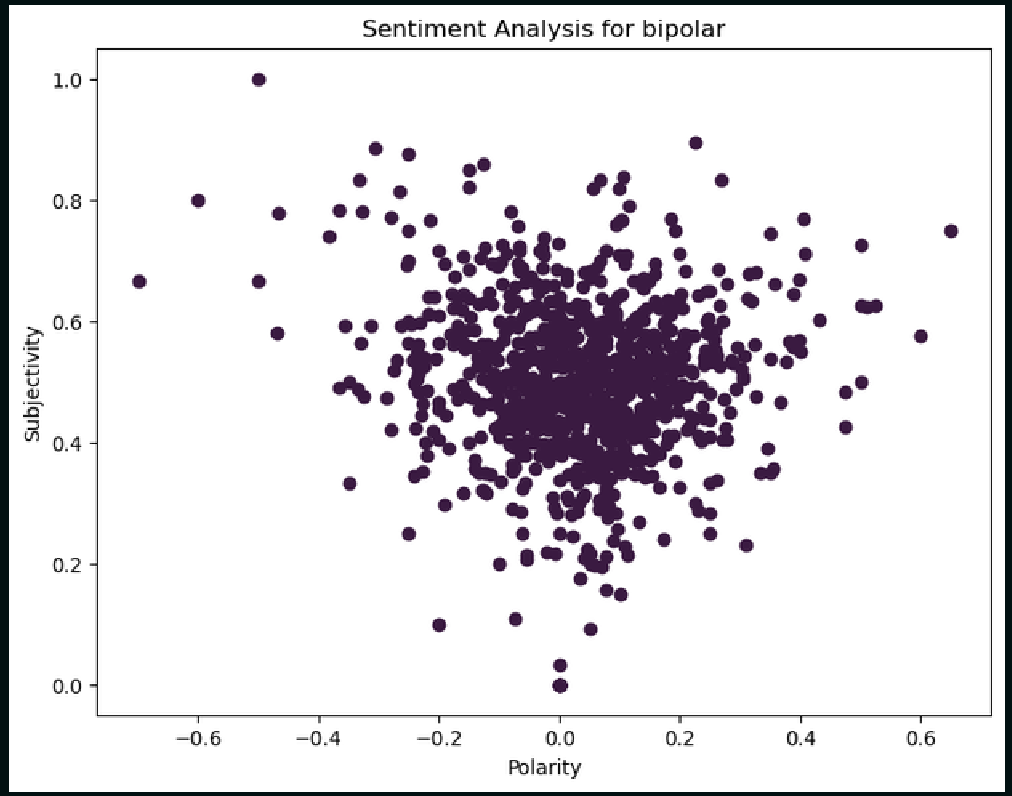
NLP
M. Health
Analysis

Schizophrenic vs Bipolar

NLP Models: TF - IDF + Gradient Boosting

Sentiment Model: TextBlob



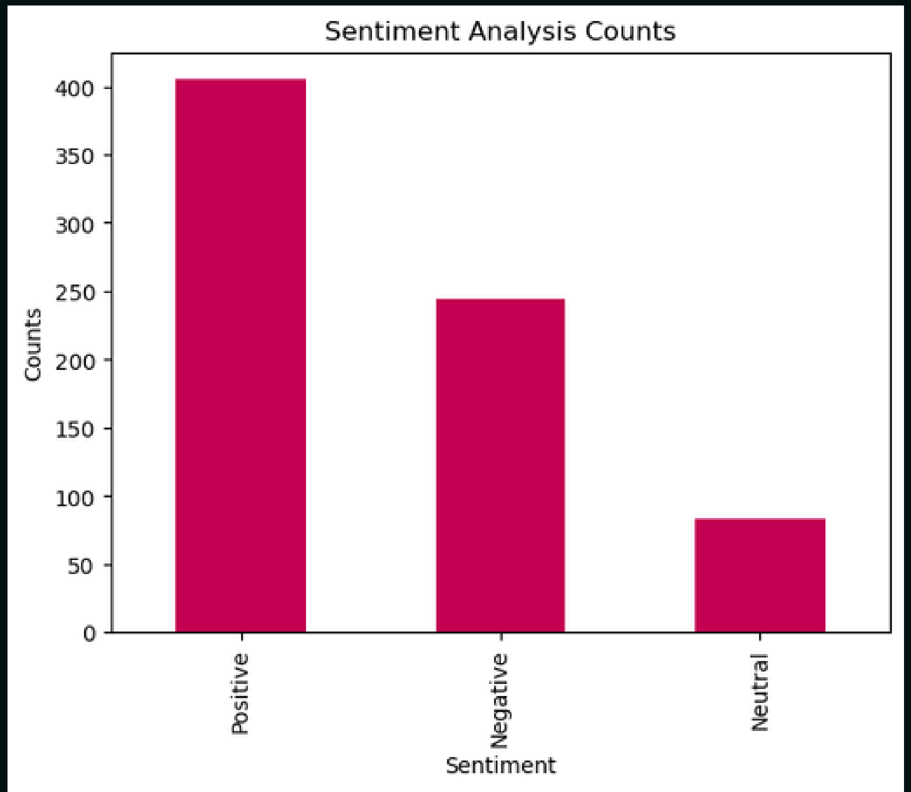


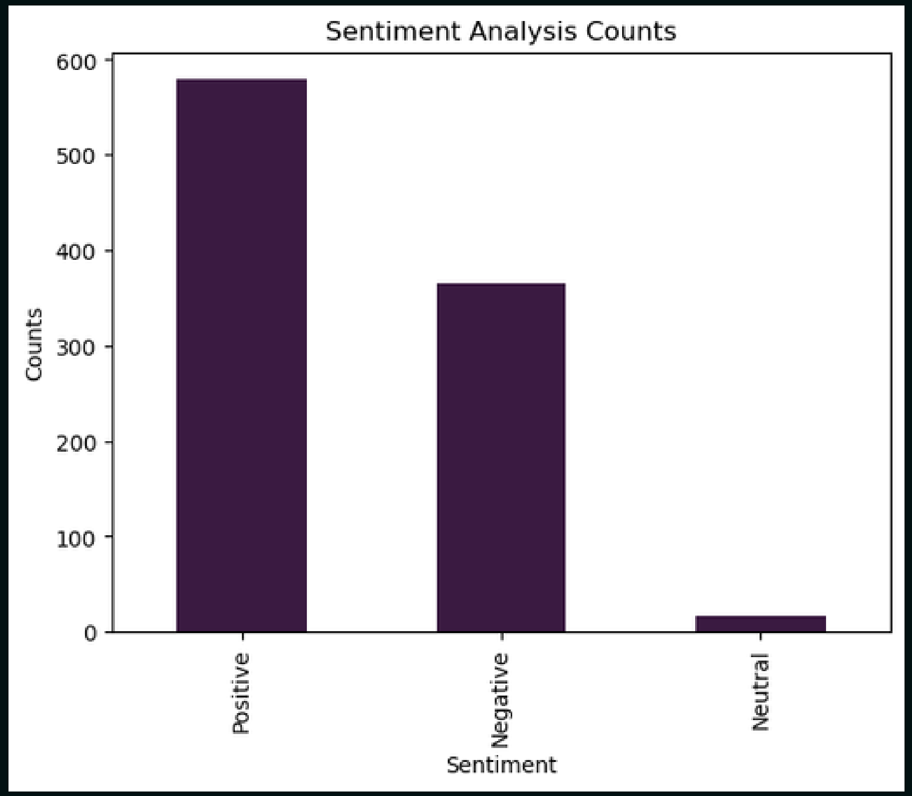
Schizophrenia

```
Positive Posts Percentage: 55.3%
Negative Posts Percentage: 33.3%
Neutral Posts Percentage: 11.3%
```

Bipolar

```
Positive Posts Percentage: 60.3%
Negative Posts Percentage: 38.0%
Neutral Posts Percentage: 1.7%
```





The Function Of Technology

Progress
80,83% +.01

Sentiments
Positive

With each addition of 200 posts to the corpus, the model become 0.01 more accurate and precise.

Future Application

-Primary: Military Leadership, Mental Health Professionals

-Secondary: HR, Business Organization

Thank you

Feel Free to ask Any Questions.

Challenges and Limitations

These are some challenges and limitations faced: Some are avoidable, and should be considered when reproducing this project; some are not and present a bottleneck for optimization

Step 01

Data Collection Timeframe
Holiday Season
[1600-2800] - raw
[400-1200] - clean

Step 02

Record-keeping measures were not implemented from the beginning

Step 03

Forum is anonymous, and no qualifying questions are asked to verify participant claim to group

Step 04

Sentiment Analysis is not able to be cross-validated automatically, as it involved understand the semantics of language

Conclusions & Recommendations

- Opt for models trained on comprehensive datasets rather than focusing solely on frequently used words.

- For sentiment analysis, consider model consistency in output before drawing conclusions.

- Recognize that aggregated sentiments might not reflect the entire community's sentiment.

- Conduct analyses outside holiday seasons to avoid seasonal influence, especially in military and mental health contexts.