Emotion Detection
Natural Language Processing - Deniz Ziştoğlu
Final Report

1. Introduction:

Problem Definition:
"How accurately can Natural Language Processing models detect emotions in social media, and how can such models be used across domains such as brand satisfaction and feedback monitoring, public response understanding, and well-being analysis of users?"
Social media platforms like Twitter (X), Instagram, and Reddit receive large amounts of emotional content constantly.
How can this emotional content be used?
• **For Businesses:** Monitor reputation in the eyes of the public and achieve real life product feedback constantly.
• **For Media Analysts**: Understand emotional responces to movies, news, or big events.
• **For Well-being of the users:** Detect potential emotional negativity in user content to give them encouraging mental health support.
• **For Researchers:** See patterns in public mood and social sentiment.

Literature Review
I have seen that foundational work such as the NRC Emotion Lexicon (Mohammad & Turney, 2013) is a seminal resource for word emotion association shows that emotions can be modeled at the word level but It can be limited by vocabulary and context. Later datasets like GoEmotions (Demszky et al., 2020) that includes reddit comments introduces a shift toward fine grained classification and multi label annotations with 27 distinct emotional categories.
WASSA (Abdul-Mageed & Ungar, 2017) which was made of tweet like sentences introduces the concept of emotion intensity in context of language processing and addressed the unique linguistic nuances of social media text.
Also:
• Traditional Approaches: Lexicon-based methods are limited by vocabulary and context.
• Modern Approaches: Transformer-based models like BERT, RoBERTa, and DistilBERT might outperform older models by understanding context and emotional nuance.
• Potential Gaps in Research:
o Models trained on some data often struggle with real world noisy, often sarcastic, slang used social media text.
o Emotion overlap and sarcasm still remains challenging.

Modern Natural Language Processing has developed from lexicon-based systems to context aware transformer models. These studies show strong results on some datasets but limited real world adaptability.

2. Dataset Selection/Comparison:

	Demszky et al. (2020) - GoEmotions (Google Research)	**Mohammad & Turney (2013) - NRC Word-Emotion Association Lexicon**	**Go Abdul-Mageed & Ungar (2017) - WASSA Emotion Intensity Dataset**
Size	211,225 Reddit comments text samples	~14,000 English words	~7,000 tweets
Label Amount	27 emotions and neutral	8 basic emotions and positive/negative sentiment	4 emotions, each rated between 0 and 1 for intensity
Example	"I can't believe they canceled my favorite show!" → *disappointment.*	"love" → joy, trust	*"I'm absolutely terrified of spiders." → Fear = 0.95*

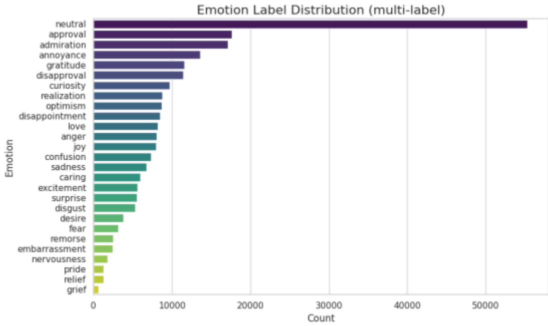
3. Pipelines & Error Analysis:

3.1 GoEmotions Dataset:

3.1.1 Dataset Overview

- **Source:** GoEmotions dataset (Demszky et al., 2020) by Google Research.
- **Size:** 211,225 reddit comments text samples.
  - Train: 168,980
  - Test: 42,245
- **Labels:** 27 emotion classes (multi label); joy, sadness, anger, fear, admiration, neutral.
- **Class Distribution:** Imbalanced.

- Common emotions include; neutral, joy, approval.
- Rare emotions include; grief, relief.



- **Average polarity distribution:** Mostly neutral.

3.1.2. Data Processing & Feature Engineering

- **Text Cleaning:**
  Removed HTML tags and other entities. Standardized spacing and characters.
- **Feature Extraction**
- **Label Transformation**
  MultiLabelBinarizer to convert emotion lists into binary vectors for multi label classification.

3.1.3. Modeling Pipeline

- **Baseline Pipeline:**
  TF IDF Vectorizer -> OneVsRest Logistic Regression.

```
...
Train size: 168980 Test size: 42245

Baseline trained in 51.0s

===== BASELINE LOGISTIC REGRESSION =====
F1-Micro: 0.23792113715574387
F1-Macro: 0.15240868360200696
```

- **Classifiers Comparison:**
  - Logistic Regression (LR)
  - Linear SVM
  - Multinomial Naive Bayes

```
===== CLASSIFIER COMPARISON =====
Logistic Regression: F1-micro=0.2379, F1-macro=0.1524, time=52.3s
Linear SVM: F1-micro=0.3194, F1-macro=0.2379, time=102.7s
Multinomial NB: F1-micro=0.0923, F1-macro=0.0467, time=8.3s

Classifier comparison summary:
                 Model  F1-Micro  F1-Macro      Time_s
0           Linear SVM  0.319407  0.237864  102.744140
1  Logistic Regression  0.237921  0.152409   52.332722
2       Multinomial NB  0.092312  0.046666    8.327369
```

- **Hyperparameter Optimization (HPO):**
  Optimized models: Logistic Regression & LinearSVC.

```
...
===== HYPERPARAMETER OPTIMIZATION (HPO) =====
Fitting 3 folds for each of 10 candidates, totalling 30 fits

LogReg HPO took 222.6s
Best params (Logistic Regression): {'clf__estimator__C': np.float64(0.8471801418819978), 'tfidf__max_features': 10000, 'tfidf__ngram_range': (1, 1)}

===== HPO Logistic Regression results =====
F1-Micro: 0.21078806426931906
F1-Macro: 0.1478113019940074
```

```
...
===== Running HPO for LinearSVC... =====
Fitting 2 folds for each of 8 candidates, totalling 16 fits

LinearSVC HPO took 228.0s
Best params (LinearSVC): {'clf__estimator__C': np.float64(2.1368329072358767), 'tfidf__max_features': 30000, 'tfidf__ngram_range': (1, 2)}

===== HPO LinearSVC results =====
F1-Micro: 0.3123741197571338
F1-Macro: 0.23401827413786477
```

3.1.4. Error Analysis

- **Per Class Weaknesses (F1 score Lowest 10 Classes)**
  - Commonly misclassified emotions: grief, relief, realization, annoyance, disappointment, embarrassment, nervousness, disapproval, pride, approval.
  - LinearSVC performs best on excitement and confusion relatively to baseline and Logistic Regression.

===== PER-LABEL METRICS – Baseline LogReg =====

	emotion	precision	recall	f1	support
15	grief	0.000000	0.000000	0.000000	115
22	relief	0.000000	0.000000	0.000000	257
21	realization	0.608696	0.008163	0.016110	1715
2	annoyance	0.333333	0.009184	0.017876	2722
8	disappointment	0.629630	0.010174	0.020024	1671
11	embarrassment	0.583333	0.013944	0.027237	502
18	nervousness	0.750000	0.016438	0.032172	365
9	disapproval	0.504950	0.022280	0.042678	2289
20	pride	0.666667	0.023077	0.044610	260
3	approval	0.604061	0.034816	0.065837	3418

===== PER-LABEL METRICS – HPO LogisticRegression =====

	emotion	precision	recall	f1	support
22	relief	0.000000	0.000000	0.000000	257
9	disapproval	0.400000	0.007864	0.015424	2289
21	realization	0.629630	0.009913	0.019518	1715
8	disappointment	0.487179	0.011370	0.022222	1671
15	grief	0.500000	0.017391	0.033613	115
2	annoyance	0.361842	0.020206	0.038274	2722
18	nervousness	0.375000	0.024658	0.046272	365
20	pride	0.777778	0.026923	0.052045	260
3	approval	0.562130	0.027794	0.052969	3418
4	caring	0.459459	0.029643	0.055692	1147

===== PER-LABEL METRICS – HPO LinearSVC =====

	emotion	precision	recall	f1	support
8	disappointment	0.276873	0.050868	0.085945	1671
21	realization	0.316176	0.050146	0.086563	1715
22	relief	0.333333	0.054475	0.093645	257
18	nervousness	0.304348	0.057534	0.096774	365
9	disapproval	0.328798	0.063346	0.106227	2289
15	grief	0.250000	0.069565	0.108844	115
2	annoyance	0.327526	0.069067	0.114078	2722
3	approval	0.367223	0.070802	0.118715	3418
12	excitement	0.388000	0.086530	0.141503	1121
5	confusion	0.396450	0.091593	0.148806	1463

- **Row Level Insights**
  Texts expressing complex or mixed emotions are the hardest to classify correctly.

- **Label Confusions & Total Misclassifed Samples:**

==================== ERROR ANALYSIS: Baseline Logistic Regression ====================

	accuracy	f1_micro	f1_macro
0	0.16802	0.237921	0.152409

Worst 10 classes by F1:

	class	precision	recall	f1	support
15	grief	0.000000	0.000000	0.000000	115
22	relief	0.000000	0.000000	0.000000	257
21	realization	0.608696	0.008163	0.016110	1715
2	annoyance	0.333333	0.009184	0.017876	2722
8	disappointment	0.629630	0.010174	0.020024	1671
11	embarrassment	0.583333	0.013944	0.027237	502
18	nervousness	0.750000	0.016438	0.032172	365
9	disapproval	0.504950	0.022280	0.042678	2289
20	pride	0.666667	0.023077	0.044610	260
3	approval	0.604061	0.034816	0.065837	3418

Total misclassified test rows: 35147 / 42245

==================== ERROR ANALYSIS: HPO Logistic Regression ====================

	accuracy	f1_micro	f1_macro
0	0.147899	0.210788	0.147811

Worst 10 classes by F1:

	class	precision	recall	f1	support
22	relief	0.000000	0.000000	0.000000	257
9	disapproval	0.400000	0.007864	0.015424	2289
21	realization	0.629630	0.009913	0.019518	1715
8	disappointment	0.487179	0.011370	0.022222	1671
15	grief	0.500000	0.017391	0.033613	115
2	annoyance	0.361842	0.020206	0.038274	2722
18	nervousness	0.375000	0.024658	0.046272	365
20	pride	0.777778	0.026923	0.052045	260
3	approval	0.562130	0.027794	0.052969	3418
4	caring	0.459459	0.029643	0.055692	1147

Total misclassified test rows: 35997 / 42245

==================== ERROR ANALYSIS: HPO Linear SVM ====================

	accuracy	f1_micro	f1_macro
0	0.222985	0.312374	0.234018

Worst 10 classes by F1:

	class	precision	recall	f1	support
8	disappointment	0.276873	0.050868	0.085945	1671
21	realization	0.316176	0.050146	0.086563	1715
22	relief	0.333333	0.054475	0.093645	257
18	nervousness	0.304348	0.057534	0.096774	365
9	disapproval	0.328798	0.063346	0.106227	2289
15	grief	0.250000	0.069565	0.108844	115
2	annoyance	0.327526	0.069067	0.114078	2722
3	approval	0.367223	0.070802	0.118715	3418
12	excitement	0.388000	0.086530	0.141503	1121
5	confusion	0.396450	0.091593	0.148806	1463

Total misclassified test rows: 32825 / 42245

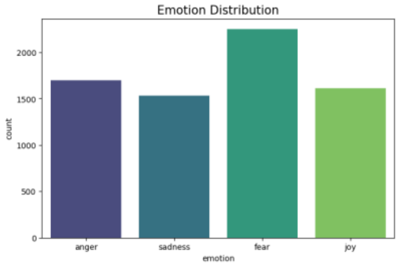
===== FINAL SUMMARY =====

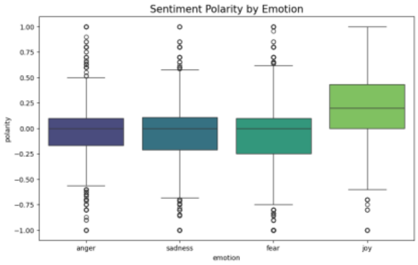
	Model	F1-Micro	F1-Macro
0	HPO LinearSVC	0.312374	0.234018
1	Baseline LogisticRegression	0.237921	0.152409
2	HPO LogisticRegression	0.210788	0.147811

3.2. WASSA Emotion Intensity Dataset

3.2.1. Dataset Overview

- **Source:** WASSA 2017 Emotion Intensity in Tweets (Go Abdul-Mageed & Ungar, 2017).
- **Dataset Composition** Four separate emotion datasets.
  Total samples: 7,102
- **Data Characteristics**
  Tweets include slang, emojis, hashtags typical to the social media.
- **Emotion distribution & Polarity Distribution**





3.2.2. Data Processing & Feature Engineering

- **Text Preprocessing**
- **Feature Engineering**

3.2.3. Modeling Pipeline

- **Baseline Pipeline**
  TF-IDF Vectorizer -> DummyClassifier

```
===== BASELINE PIPELINE =====
Accuracy: 0.31738212526389864
F1_macro: 0.12045940170940171

Classification Report:
              precision    recall  f1-score   support

       anger       0.00      0.00      0.00       340
        fear       0.32      1.00      0.48       451
         joy       0.00      0.00      0.00       323
     sadness       0.00      0.00      0.00       307

    accuracy                           0.32      1421
   macro avg       0.08      0.25      0.12      1421
weighted avg       0.10      0.32      0.15      1421
```

- **Classifier Comparison**
  Models Tested;
  - Logistic Regression
  - Linear SVM (LinearSVC)

```
Training LogisticRegression...


LogisticRegression Classification Report:
              precision    recall  f1-score   support

       anger       0.87      0.83      0.85       340
        fear       0.80      0.91      0.85       451
         joy       0.92      0.87      0.89       323
     sadness       0.84      0.77      0.80       307

    accuracy                           0.85      1421
   macro avg       0.86      0.84      0.85      1421
weighted avg       0.85      0.85      0.85      1421
```

```
Training LinearSVC...


LinearSVC Classification Report:
              precision    recall  f1-score   support

       anger       0.89      0.89      0.89       340
        fear       0.86      0.91      0.89       451
         joy       0.95      0.92      0.94       323
     sadness       0.85      0.82      0.84       307

    accuracy                           0.89      1421
   macro avg       0.89      0.88      0.89      1421
weighted avg       0.89      0.89      0.89      1421
```

- **Hyperparameter Optimization (HPO):**
  Optimized models: Logistic Regression & LinearSVC

```
Running HPO for LogisticRegression...

Fitting 3 folds for each of 20 candidates, totalling 60 fits

Running HPO for LinearSVC...

Fitting 3 folds for each of 20 candidates, totalling 60 fits

===== BEST LOGISTIC REGRESSION MODEL =====
{'clf__C': np.float64(5.242344384136116), 'clf__penalty': 'l2'
              precision    recall  f1-score   support

       anger       0.88      0.86      0.87       340
        fear       0.84      0.91      0.87       451
         joy       0.95      0.91      0.93       323
     sadness       0.85      0.80      0.82       307

    accuracy                           0.87      1421
   macro avg       0.88      0.87      0.87      1421
weighted avg       0.88      0.87      0.87      1421
```

```
===== BEST LINEAR SVC MODEL =====
{'clf__C': np.float64(3.845401188473625), 'tfidf__min_df': 1,
              precision    recall  f1-score   support

       anger       0.88      0.88      0.88       340
        fear       0.87      0.90      0.88       451
         joy       0.95      0.92      0.93       323
     sadness       0.84      0.82      0.83       307

    accuracy                           0.88      1421
   macro avg       0.88      0.88      0.88      1421
weighted avg       0.88      0.88      0.88      1421
```

3.2.4. Error Analysis

- **Most Confused Emotion Pairs**

```
===== MOST CONFUSED EMOTION PAIRS – Baseline Dummy =====
True='anger' → Predicted='fear' : 340
True='joy' → Predicted='fear' : 323
True='sadness' → Predicted='fear' : 307

===== MOST CONFUSED EMOTION PAIRS – HPO LogisticRegression =====
True='sadness' → Predicted='fear' : 38
True='anger' → Predicted='fear' : 24
True='fear' → Predicted='sadness' : 22
True='sadness' → Predicted='anger' : 18
True='anger' → Predicted='sadness' : 17
True='joy' → Predicted='fear' : 17
True='fear' → Predicted='anger' : 15
True='joy' → Predicted='anger' : 8
True='sadness' → Predicted='joy' : 6
True='anger' → Predicted='joy' : 5

===== MOST CONFUSED EMOTION PAIRS – HPO LinearSVC =====
True='sadness' → Predicted='fear' : 29
True='fear' → Predicted='sadness' : 24
True='anger' → Predicted='fear' : 18
True='anger' → Predicted='sadness' : 18
True='fear' → Predicted='anger' : 18
True='sadness' → Predicted='anger' : 18
True='joy' → Predicted='fear' : 14
True='sadness' → Predicted='joy' : 7
True='joy' → Predicted='anger' : 6
True='anger' → Predicted='joy' : 5
```

- **Key Failure Modes**
  - Sadness vs. Fear are frequently confused.
  - Anger vs. Sadness appears when tweets express frustration without explicit aggression.
  - Joy misclassified as Fear usually involves sarcasm, irony, or ambiguous emojis.

===== PER-CLASS METRICS – Baseline Dummy =====

	emotion	precision	recall	f1	support
0	anger	0.000000	0.0	0.000000	340
2	joy	0.000000	0.0	0.000000	323
3	sadness	0.000000	0.0	0.000000	307
1	fear	0.317382	1.0	0.481838	451

===== PER-CLASS METRICS – HPO LogisticRegression

	emotion	precision	recall	f1	support
3	sadness	0.850694	0.798046	0.823529	307
0	anger	0.877612	0.864706	0.871111	340
1	fear	0.838115	0.906874	0.871140	451
2	joy	0.948387	0.910217	0.928910	323

===== PER-CLASS METRICS – HPO LinearSVC =====

	emotion	precision	recall	f1	support
3	sadness	0.843333	0.824104	0.833608	307
0	anger	0.876833	0.879412	0.878120	340
1	fear	0.868817	0.895787	0.882096	451
2	joy	0.946032	0.922601	0.934169	323

===== FINAL PERFORMANCE SUMMARY =====

	Model	F1-Micro	F1-Macro
0	HPO LinearSVC	0.882477	0.881998
1	HPO LogisticRegression	0.874032	0.873672
2	Baseline Dummy	0.317382	0.120459

- **Row Level Insights**
  HPO models reduce errors by >70% compared to baseline.
  Misclassifications often involves:
  - sarcastic or ironic tone
  - mixed emotional states
  - slang

===== ERROR TABLE – Baseline Dummy =====
Total misclassified samples: 970 / 1421

	index	text	true	pred
0	0	a good thing about being sick is that coughing...	joy	fear
1	1	mhchat childhood experiences inform adult rela...	sadness	fear
2	3	colinkaepernick protests because of the unjust...	anger	fear
3	6	they unleashing gods wrath just as cain did	anger	fear
4	7	oh daaaaaaamn getitgirl girlpower rawr chica...	anger	fear
5	8	extreme sadness	sadness	fear
6	10	meogysdailyvocabenhancment elation quality or ...	joy	fear
7	11	why doesnt anybody i know watch penny dreadful	sadness	fear
8	12	whatever you decide to do make sure it makes y...	joy	fear
9	13	that feel when you travel 700 miles to pick up...	anger	fear
10	14	watch this amazing lively broadcast by music...	joy	fear
11	15	and i will strike down upon thee with great ve...	anger	fear
12	17	one chosen by the clp members mp seats are not...	anger	fear
13	18	has anyone noticed that stories in recent day...	sadness	fear
14	20	sadly i have to go to bed now dont announce an...	sadness	fear
15	22	literally being here makes me depress tbh	sadness	fear
16	23	hi little late to lost girl but think you awes...	sadness	fear
17	24	then i look at you and we burst out laughing	anger	fear
18	25	can we get a shot of lingys face at 14 time p...	anger	fear

===== ERROR TABLE – HPO LogisticRegression =====
Total misclassified samples: 179 / 1421

	index	text	true	pred
0	11	why doesnt anybody i know watch penny dreadful	sadness	fear
1	25	can we get a shot of lingys face at 14 time p...	anger	fear
2	33	my haters are like crickets they chirp all day...	joy	anger
3	34	quite simply the worst airline worstairline iv...	sadness	fear
4	38	lmao i can only imagine the frown across that ...	sadness	anger
5	51	shoutout to the drunk man on the bus who pisse...	sadness	anger
6	62	heading home to cut grass in the heat all i wa...	anger	sadness
7	80	turn that frown upside down buddy 😔	sadness	anger
8	88	sometimes i think the british political landsc...	anger	sadness
9	92	time for some despair sdr3 despair fuckthisanime	sadness	fear
10	127	today has dragged on	fear	anger
11	134	hope was an instinct only the reasoning human ...	sadness	fear
12	139	i liked that she was not moping around in all ...	sadness	joy
13	140	hes spent his campaign dividing people up and ...	anger	fear
14	146	happy birthday to stephen king a man responsib...	fear	joy
15	149	thats how you start a season thats how you ope...	sadness	fear
16	154	at work yesterday some old cunt couple told me...	anger	fear
17	160	candices pout is gonna take someone eye out ma...	sadness	anger
18	165	my alarm clock was ringing this morning n my f...	fear	anger

===== ERROR TABLE – HPO LinearSVC =====
Total misclassified samples: 167 / 1421

	index	text	true	pred
0	11	why doesnt anybody i know watch penny dreadful	sadness	fear
1	25	can we get a shot of lingys face at 14 time p...	anger	fear
2	34	quite simply the worst airline worstairline iv...	sadness	fear
3	38	lmao i can only imagine the frown across that ...	sadness	anger
4	51	shoutout to the drunk man on the bus who pisse...	sadness	anger
5	62	heading home to cut grass in the heat all i wa...	anger	sadness
6	80	turn that frown upside down buddy 😔	sadness	anger
7	88	sometimes i think the british political landsc...	anger	sadness
8	92	time for some despair sdr3 despair fuckthisanime	sadness	fear
9	94	whats up cowboys	fear	joy
10	127	today has dragged on	fear	anger
11	134	hope was an instinct only the reasoning human ...	sadness	fear
12	139	i liked that she was not moping around in all ...	sadness	joy
13	140	hes spent his campaign dividing people up and ...	anger	fear
14	146	happy birthday to stephen king a man responsib...	fear	joy
15	149	thats how you start a season thats how you ope...	sadness	fear
16	160	candices pout is gonna take someone eye out ma...	sadness	anger
17	165	my alarm clock was ringing this morning n my f...	fear	anger
18	194	wont use using again these guys cant get not...	anger	fear

3.2.5. Transformer Model

Bert Transformer Model is pre-trained language model that utilizes the Transformer architecture to generate high-quality representations of text.

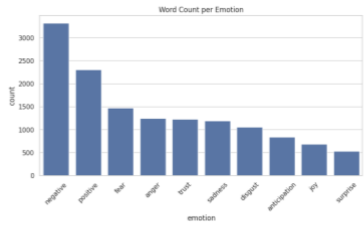
```
Epoch 1: 100%|██████████| 356/356 [50:06<00:00,  8.45s/it, loss=0.0567]
Epoch 1 average loss: 0.8521
Epoch 2: 100%|██████████| 356/356 [48:38<00:00,  8.20s/it, loss=0.0382]
Epoch 2 average loss: 0.3164
Epoch 3: 100%|██████████| 356/356 [48:18<00:00,  8.14s/it, loss=0.0293]Epoch 3 average loss: 0.1967
```

```
===== BERT RESULTS =====
Accuracy: 0.8733286418015482
Macro F1: 0.8746920812128666

Classification Report:
              precision    recall  f1-score   support

       anger       0.93      0.79      0.86       340
        fear       0.83      0.90      0.86       451
         joy       0.95      0.91      0.93       323
     sadness       0.83      0.88      0.85       307

    accuracy                           0.87      1421
   macro avg       0.88      0.87      0.87      1421
weighted avg       0.88      0.87      0.87      1421
```

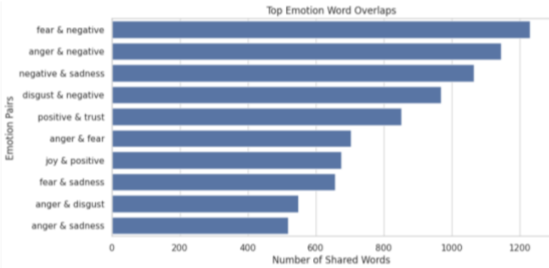
3.3. NRC Word Emotion Lexicon

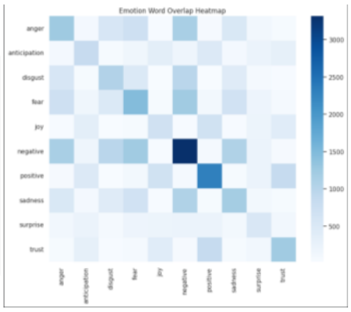
3.3.1. Dataset Overview

- **Source:** NRC Word–Emotion Association Lexicon (Mohammad & Turney, 2013)
- **Size:** Total extracted unique words = **13,800**
- **Labels:** Contains words with 10 emotion labels:
  anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise, trust
- **Class Distribution:** is highly imbalanced.



- **Key Dataset Problems**
  - **Single word classification**: Extremely sparse textual information.
  - **Emotion overlap**: Many words occur in multiple emotions.





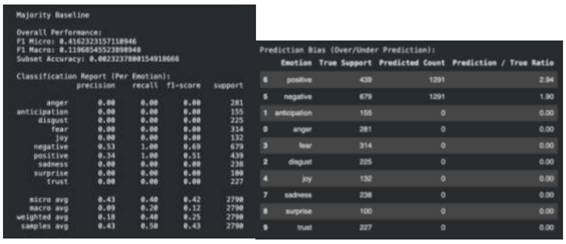
- o **Short text feature limitation**: Some models struggle to capture semantics from just one word.

3.3. 2. Data Processing & Feature Engineering

- **Text Cleaning**
- **Feature Extraction**
- **Train/Test Split**

3.3. 3. Modeling Pipeline

- **Baseline Methods**
  - o **Majority Class Baseline**

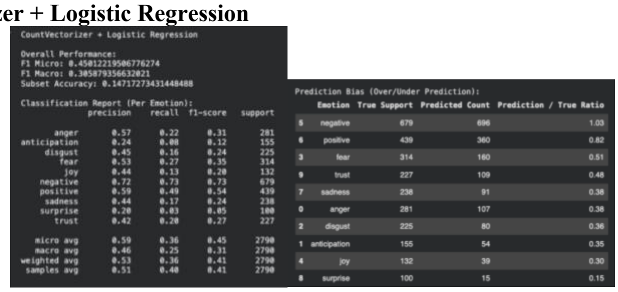


```
Majority Baseline

Overall Performance:
F1 Micro: 0.4162323157110946
F1 Macro: 0.11968545523898948
Subset Accuracy: 0.0023237800154918666

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.00      0.00      0.00       281
anticipation       0.00      0.00      0.00       155
     disgust       0.00      0.00      0.00       225
        fear       0.00      0.00      0.00       314
         joy       0.00      0.00      0.00       132
    negative       0.53      1.00      0.69       679
    positive       0.34      1.00      0.51       439
     sadness       0.00      0.00      0.00       238
    surprise       0.00      0.00      0.00       100
       trust       0.00      0.00      0.00       227

   micro avg       0.43      0.40      0.42      2790
   macro avg       0.09      0.20      0.12      2790
weighted avg       0.18      0.40      0.25      2790
 samples avg       0.43      0.50      0.43      2790
```

Prediction Bias (Over/Under Prediction):

	Emotion	True Support	Predicted Count	Prediction / True Ratio
6	positive	439	1291	2.94
5	negative	679	1291	1.90
1	anticipation	155	0	0.00
0	anger	281	0	0.00
3	fear	314	0	0.00
2	disgust	225	0	0.00
4	joy	132	0	0.00
7	sadness	238	0	0.00
8	surprise	100	0	0.00
9	trust	227	0	0.00

  - o **CountVectorizer + Logistic Regression**

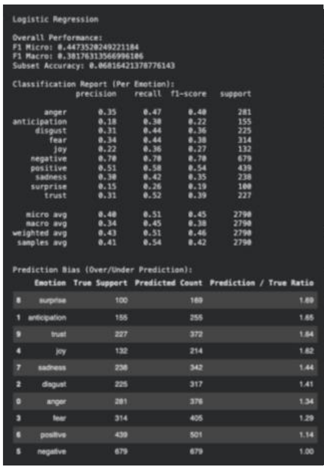


```
CountVectorizer + Logistic Regression

Overall Performance:
F1 Micro: 0.45012219506776274
F1 Macro: 0.3058793566320 21
Subset Accuracy: 0.14717273431448488

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.57      0.22      0.31       281
anticipation       0.24      0.08      0.12       155
     disgust       0.45      0.16      0.24       225
        fear       0.53      0.27      0.35       314
         joy       0.44      0.13      0.20       132
    negative       0.72      0.73      0.73       679
    positive       0.59      0.49      0.54       439
     sadness       0.44      0.17      0.24       238
    surprise       0.20      0.03      0.05       100
       trust       0.42      0.20      0.27       227

   micro avg       0.59      0.36      0.45      2790
   macro avg       0.46      0.25      0.31      2790
weighted avg       0.53      0.36      0.41      2790
 samples avg       0.51      0.40      0.41      2790
```

Prediction Bias (Over/Under Prediction):

	Emotion	True Support	Predicted Count	Prediction / True Ratio
5	negative	679	696	1.03
6	positive	439	360	0.82
3	fear	314	160	0.51
9	trust	227	109	0.48
7	sadness	238	91	0.38
0	anger	281	107	0.38
2	disgust	225	80	0.36
1	anticipation	155	54	0.35
4	joy	132	39	0.30
8	surprise	100	15	0.15

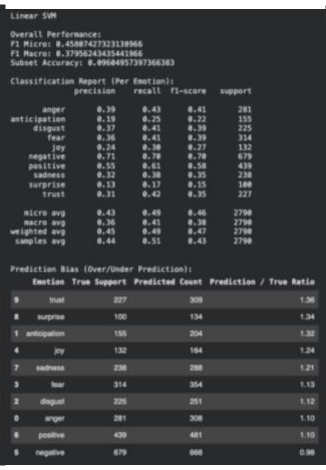
- **Classifier Comparison**
  - o Logistic Regression
  - o Linear SVM
  - o Multinomial NB



```
Logistic Regression

Overall Performance:
F1 Micro: 0.4473520249221184
F1 Macro: 0.38176313566996106
Subset Accuracy: 0.06816421378776143

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.35      0.47      0.40       281
anticipation       0.18      0.30      0.22       155
     disgust       0.31      0.44      0.36       225
        fear       0.34      0.44      0.38       314
         joy       0.22      0.36      0.27       132
    negative       0.70      0.70      0.70       679
    positive       0.51      0.58      0.54       439
     sadness       0.30      0.42      0.35       238
    surprise       0.15      0.26      0.19       100
       trust       0.31      0.52      0.39       227

   micro avg       0.40      0.51      0.45      2790
   macro avg       0.34      0.45      0.38      2790
weighted avg       0.43      0.51      0.46      2790
 samples avg       0.41      0.54      0.42      2790
```

Prediction Bias (Over/Under Prediction):

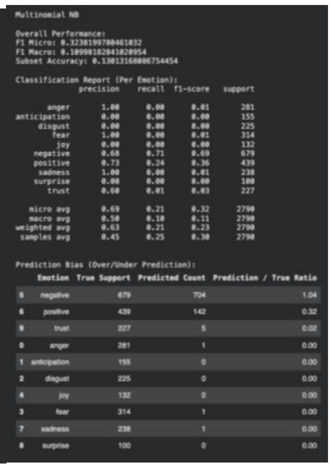
	Emotion	True Support	Predicted Count	Prediction / True Ratio
8	surprise	100	169	1.69
1	anticipation	155	255	1.65
9	trust	227	372	1.64
4	joy	132	214	1.62
7	sadness	238	342	1.44
2	disgust	225	317	1.41
0	anger	281	376	1.34
3	fear	314	405	1.29
6	positive	439	501	1.14
5	negative	679	679	1.00



```
Linear SVM

Overall Performance:
F1 Micro: 0.45807427323138966
F1 Macro: 0.37956243435441966
Subset Accuracy: 0.09604957397366383

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.39      0.43      0.41       281
anticipation       0.19      0.25      0.22       155
     disgust       0.37      0.41      0.39       225
        fear       0.36      0.41      0.39       314
         joy       0.24      0.30      0.27       132
    negative       0.71      0.70      0.70       679
    positive       0.55      0.61      0.58       439
     sadness       0.32      0.38      0.35       238
    surprise       0.13      0.17      0.15       100
       trust       0.31      0.42      0.35       227

   micro avg       0.43      0.49      0.46      2790
   macro avg       0.36      0.41      0.38      2790
weighted avg       0.45      0.49      0.47      2790
 samples avg       0.44      0.51      0.43      2790
```

Prediction Bias (Over/Under Prediction):

	Emotion	True Support	Predicted Count	Prediction / True Ratio
9	trust	227	309	1.36
8	surprise	100	134	1.34
1	anticipation	155	204	1.32
4	joy	132	164	1.24
7	sadness	238	288	1.21
3	fear	314	354	1.13
2	disgust	225	251	1.12
0	anger	281	308	1.10
6	positive	439	481	1.10
5	negative	679	668	0.98



```
Multinomial NB

Overall Performance:
F1 Micro: 0.3238199780461032
F1 Macro: 0.10998182841020954
Subset Accuracy: 0.13013168086754454

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       1.00      0.00      0.01       281
anticipation       0.00      0.00      0.00       155
     disgust       0.00      0.00      0.00       225
        fear       1.00      0.00      0.01       314
         joy       0.00      0.00      0.00       132
    negative       0.68      0.71      0.69       679
    positive       0.73      0.24      0.36       439
     sadness       1.00      0.00      0.01       238
    surprise       0.00      0.00      0.00       100
       trust       0.60      0.01      0.03       227

   micro avg       0.69      0.21      0.32      2790
   macro avg       0.50      0.10      0.11      2790
weighted avg       0.63      0.21      0.23      2790
 samples avg       0.45      0.25      0.30      2790
```

Prediction Bias (Over/Under Prediction):

	Emotion	True Support	Predicted Count	Prediction / True Ratio
5	negative	679	704	1.04
6	positive	439	142	0.32
9	trust	227	5	0.02
0	anger	281	1	0.00
1	anticipation	155	0	0.00
2	disgust	225	0	0.00
4	joy	132	0	0.00
3	fear	314	1	0.00
7	sadness	238	1	0.00
8	surprise	100	0	0.00

- **Hyperparameter Optimization (HPO)**
  - o Logistic Regression
  - o LinearSVM

```
HPO Logistic Regression

Overall Performance:
F1 Micro: 0.44562243502051985
F1 Macro: 0.36935504083342957
Subset Accuracy: 0.09604957397366383

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.38      0.37      0.37       281
anticipation       0.16      0.20      0.18       155
     disgust       0.35      0.40      0.37       225
        fear       0.39      0.41      0.40       314
         joy       0.27      0.34      0.30       132
    negative       0.69      0.68      0.69       679
    positive       0.53      0.57      0.55       439
     sadness       0.32      0.37      0.34       238
    surprise       0.15      0.18      0.16       100
       trust       0.29      0.39      0.33       227

   micro avg       0.43      0.47      0.45      2790
   macro avg       0.35      0.39      0.37      2790
weighted avg       0.44      0.47      0.45      2790
 samples avg       0.42      0.50      0.41      2790

Prediction Bias (Over/Under Prediction):
```

	Emotion	True Support	Predicted Count	Prediction / True Ratio
9	trust	227	303	1.33
4	joy	132	167	1.27
1	anticipation	155	195	1.26
8	surprise	100	121	1.21
7	sadness	238	272	1.14
2	disgust	225	252	1.12
6	positive	439	471	1.07
3	fear	314	332	1.06
5	negative	679	672	0.99
0	anger	281	273	0.97

```
HPO Linear SVM

Overall Performance:
F1 Micro: 0.455749354005168
F1 Macro: 0.38161711410234667
Subset Accuracy: 0.07823392718822618

Classification Report (Per Emotion):
              precision    recall  f1-score   support

       anger       0.39      0.46      0.42       281
anticipation       0.18      0.27      0.22       155
     disgust       0.33      0.43      0.37       225
        fear       0.36      0.43      0.39       314
         joy       0.22      0.31      0.26       132
    negative       0.71      0.71      0.71       679
    positive       0.54      0.60      0.57       439
     sadness       0.30      0.42      0.35       238
    surprise       0.14      0.21      0.17       100
       trust       0.30      0.45      0.36       227

   micro avg       0.41      0.51      0.46      2790
   macro avg       0.35      0.43      0.38      2790
weighted avg       0.44      0.51      0.47      2790
 samples avg       0.42      0.53      0.42      2790

Prediction Bias (Over/Under Prediction):
```

	Emotion	True Support	Predicted Count	Prediction / True Ratio
8	surprise	100	154	1.54
9	trust	227	344	1.52
1	anticipation	155	231	1.49
4	joy	132	186	1.41
7	sadness	238	327	1.37
2	disgust	225	289	1.28
0	anger	281	334	1.19
3	fear	314	374	1.19
6	positive	439	489	1.11
5	negative	679	674	0.99

3.3.4. Error Analysis

- **Worst emotions per model**
  - Lowest performance observed on low-frequency emotions.
  - Commonly weak labels are surprise, anticipation, trust.
  - Errors are driven by emotion overlap and limited semantic information from single words.

Worst emotions – Majority Baseline

	Emotion	Precision	Recall	F1	Support
0	anger	0.0	0.0	0.0	281
1	anticipation	0.0	0.0	0.0	155
2	disgust	0.0	0.0	0.0	225
3	fear	0.0	0.0	0.0	314
4	joy	0.0	0.0	0.0	132

Worst emotions – CountVectorizer Baseline

	Emotion	Precision	Recall	F1	Support
8	surprise	0.200000	0.030000	0.052174	100
1	anticipation	0.240741	0.083871	0.124402	155
4	joy	0.435897	0.128788	0.198830	132
2	disgust	0.450000	0.160000	0.236066	225
7	sadness	0.439560	0.168067	0.243161	238

Worst emotions – Logistic Regression

	Emotion	Precision	Recall	F1	Support
8	surprise	0.153846	0.260000	0.193309	100
1	anticipation	0.180392	0.296774	0.224390	155
4	joy	0.219626	0.356061	0.271676	132
7	sadness	0.295322	0.424370	0.348276	238
2	disgust	0.309148	0.435556	0.361624	225

Worst emotions – Linear SVM

	Emotion	Precision	Recall	F1	Support
8	surprise	0.126866	0.170000	0.145299	100
1	anticipation	0.191176	0.251613	0.217270	155
4	joy	0.243902	0.303030	0.270270	132
7	sadness	0.315972	0.382353	0.346008	238
9	trust	0.307443	0.418502	0.354478	227

Worst emotions – Multinomial NB

	Emotion	Precision	Recall	F1	Support
1	anticipation	0.0	0.000000	0.000000	155
2	disgust	0.0	0.000000	0.000000	225
4	joy	0.0	0.000000	0.000000	132
8	surprise	0.0	0.000000	0.000000	100
3	fear	1.0	0.003185	0.006349	314

Worst emotions – HPO Logistic Regression

	Emotion	Precision	Recall	F1	Support
8	surprise	0.148760	0.180000	0.162896	100
1	anticipation	0.158974	0.200000	0.177143	155
4	joy	0.269461	0.340909	0.301003	132
9	trust	0.290429	0.387665	0.332075	227
7	sadness	0.319653	0.365546	0.341176	238

Worst emotions – HPO Linear SVM

	Emotion	Precision	Recall	F1	Support
8	surprise	0.136364	0.210000	0.165354	100
1	anticipation	0.181818	0.270968	0.217617	155
4	joy	0.220430	0.310606	0.257862	132
7	sadness	0.302752	0.415966	0.350442	238
9	trust	0.296512	0.449339	0.357268	227

- **False Positives & False Negatives per model**
  - High false positives for dominant labels (positive, negative).
  - High false negatives for minority emotions.
  - Class imbalance strongly affects recall.

Majority Baseline – FP vs FN

	Emotion	False Positives	False Negatives
3	fear	0	314
0	anger	0	281
7	sadness	0	238
9	trust	0	227
2	disgust	0	225
1	anticipation	0	155
4	joy	0	132
8	surprise	0	100
5	negative	612	0
6	positive	852	0

Linear SVM – FP vs FN

	Emotion	False Positives	False Negatives
5	negative	194	205
3	fear	225	185
6	positive	215	173
0	anger	188	161
7	sadness	197	147
2	disgust	159	133
9	trust	214	132
1	anticipation	165	116
4	joy	124	92
8	surprise	117	83

CountVectorizer Baseline – FP vs FN

	Emotion	False Positives	False Negatives
3	fear	76	230
6	positive	146	225
0	anger	46	220
7	sadness	51	198
2	disgust	44	189
9	trust	63	181
5	negative	197	180
1	anticipation	41	142
4	joy	22	115
8	surprise	12	97

Multinomial NB – FP vs FN

	Emotion	False Positives	False Negatives
6	positive	38	335
3	fear	0	313
0	anger	0	280
7	sadness	0	237
2	disgust	0	225
9	trust	2	224
5	negative	224	199
1	anticipation	0	155
4	joy	0	132
8	surprise	0	100

Logistic Regression – FP vs FN

	Emotion	False Positives	False Negatives
5	negative	202	202
6	positive	246	184
3	fear	267	176
0	anger	245	150
7	sadness	241	137
2	disgust	219	127
9	trust	255	110
1	anticipation	209	109
4	joy	167	85
8	surprise	143	74

HPO Logistic Regression – FP vs FN

	Emotion	False Positives	False Negatives
5	negative	209	216
6	positive	221	189
3	fear	203	185
0	anger	170	178
7	sadness	185	151
9	trust	215	139
2	disgust	163	136
1	anticipation	164	124
4	joy	122	87
8	surprise	103	82

HPO Linear SVM – FP vs FN

	Emotion	False Positives	False Negatives
5	negative	193	198
3	fear	239	179
6	positive	224	174
0	anger	205	152
7	sadness	228	139
2	disgust	193	129
9	trust	242	125
1	anticipation	189	113
4	joy	145	91
8	surprise	133	79

- **Sample error analysis**
  - Single words mapped to multiple emotions might cause ambiguity
  - Semantically similar emotions are frequently confused
  - Errors are consistent across most models

Majority Baseline – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[negative, positive]	2
1	administrative	[trust]	[negative, positive]	3
2	constraint	[anger, fear, negative, sadness]	[negative, positive]	4
3	investigation	[anticipation]	[negative, positive]	3
4	wane	[negative, sadness]	[negative, positive]	2
5	repellent	[anger, disgust, fear, negative]	[negative, positive]	4
6	denounce	[anger, disgust, negative]	[negative, positive]	3
7	clown	[anticipation, joy, positive, surprise]	[negative, positive]	4
8	mutant	[negative]	[negative, positive]	1
9	primacy	[positive]	[negative, positive]	1

CountVectorizer Baseline – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[positive]	3
1	administrative	[trust]	[negative, positive]	3
2	constraint	[anger, fear, negative, sadness]	[positive]	5
3	investigation	[anticipation]	[positive]	2
4	wane	[negative, sadness]	[fear, negative]	2
5	repellent	[anger, disgust, fear, negative]	[disgust, fear, negative]	1
6	denounce	[anger, disgust, negative]	[anger, negative]	1
7	clown	[anticipation, joy, positive, surprise]	[negative]	5
8	mutant	[negative]	[]	1
9	affiliated	[positive]	[negative]	2

Logistic Regression – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[positive]	3
1	administrative	[trust]	[negative, positive, trust]	2
2	constraint	[anger, fear, negative, sadness]	[positive, trust]	6
3	investigation	[anticipation]	[negative]	2
4	wane	[negative, sadness]	[anger, fear, negative, sadness]	2
5	repellent	[anger, disgust, fear, negative]	[disgust, fear, positive]	3
6	denounce	[anger, disgust, negative]	[anger, anticipation, fear, negative]	3
7	clown	[anticipation, joy, positive, surprise]	[anticipation, negative, sadness]	5
8	mutant	[negative]	[anger, anticipation, disgust, positive]	5
9	primacy	[positive]	[anticipation, positive, trust]	2

Linear SVM – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[positive]	3
1	administrative	[trust]	[negative, positive]	3
2	constraint	[anger, fear, negative, sadness]	[joy, positive]	6
3	investigation	[anticipation]	[positive]	2
4	wane	[negative, sadness]	[anger, fear, negative, sadness]	2
5	repellent	[anger, disgust, fear, negative]	[disgust, fear, negative, positive]	2
6	denounce	[anger, disgust, negative]	[anger, anticipation, negative]	2
7	clown	[anticipation, joy, positive, surprise]	[anticipation, negative, sadness]	5
8	mutant	[negative]	[anger, anticipation, positive]	4
9	primacy	[positive]	[anticipation, positive, trust]	2

Multinomial NB – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[]	2
1	administrative	[trust]	[negative]	2
2	constraint	[anger, fear, negative, sadness]	[]	4
3	investigation	[anticipation]	[negative]	2
4	wane	[negative, sadness]	[negative]	1
5	repellent	[anger, disgust, fear, negative]	[]	4
6	denounce	[anger, disgust, negative]	[negative]	2
7	clown	[anticipation, joy, positive, surprise]	[negative]	5
8	mutant	[negative]	[]	1
9	affiliated	[positive]	[negative]	2

HPO Logistic Regression – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[joy, positive]	4
1	administrative	[trust]	[negative]	2
2	constraint	[anger, fear, negative, sadness]	[positive]	5
3	investigation	[anticipation]	[positive]	2
4	wane	[negative, sadness]	[fear, negative, sadness]	1
5	repellent	[anger, disgust, fear, negative]	[disgust, fear, negative, positive]	2
6	denounce	[anger, disgust, negative]	[anger, anticipation, negative]	2
7	clown	[anticipation, joy, positive, surprise]	[negative, sadness]	6
8	mutant	[negative]	[anticipation, positive]	3
9	primacy	[positive]	[anger, positive, trust]	2

HPO Linear SVM – sample errors

	word	true	pred	num_errors
0	swine	[disgust, negative]	[positive]	3
1	administrative	[trust]	[negative, positive]	3
2	constraint	[anger, fear, negative, sadness]	[positive, trust]	6
3	investigation	[anticipation]	[negative]	2
4	wane	[negative, sadness]	[anger, fear, negative, sadness]	2
5	repellent	[anger, disgust, fear, negative]	[disgust, fear, negative, positive]	2
6	denounce	[anger, disgust, negative]	[anger, anticipation, fear, negative]	3
7	clown	[anticipation, joy, positive, surprise]	[anticipation, negative, sadness]	5
8	mutant	[negative]	[anger, anticipation, disgust, positive]	5
9	primacy	[positive]	[anticipation, positive, trust]	2

- **Best predicted labels & Best models for each label**
    - Best performance achieved on frequent emotions
    - Strongest labels are positive, negative, anger, joy
    - LinearSVM performs best overall
    - No model consistently handles rare emotions well

	True to Predicted	Count
0	(negative, fear)	249
1	(negative, anger)	234
2	(negative, sadness)	225
3	(fear, negative)	209
4	(negative, disgust)	202
5	(anger, negative)	199
6	(positive, trust)	172
7	(disgust, negative)	165
8	(sadness, negative)	159
9	(negative, positive)	154

- **Error Summary**
    - Majority of errors caused by:
      - class imbalance
      - emotion overlap
      - single-word input limitation
    - Optimized models reduce errors but do not eliminate ambiguity
    - Dataset characteristics dominate model performance

	Model	F1-Micro	F1-Macro	Total Errors	Error Rate
3	Linear SVM	0.458074	0.379562	3225	0.249806
6	HPO Linear SVM	0.455749	0.381617	3370	0.261038
1	CountVectorizer Baseline	0.450122	0.305879	2475	0.191712
2	Logistic Regression	0.447352	0.381763	3548	0.274826
5	HPO Logistic Regression	0.445622	0.369355	3242	0.251123
0	Majority Baseline	0.416232	0.119685	3136	0.242912
4	Multinomial NB	0.323820	0.109982	2464	0.190860

4. Discussion and Future Work

This study investigated the effectiveness of multiple Natural Language Processing approaches for emotion detection in social media text using three complementary but very different datasets: GoEmotions, WASSA Emotion Intensity, and the NRC Emotion Lexicon. The experimental results demonstrate that model performance is strongly influenced by both dataset characteristics and task formulation. Across all datasets, classical machine learning models combined with TF–IDF features provided good foundational baselines, while HPO models offered improvements but did not fully resolve challenges related to emotion overlapping and class imbalance.

On the GoEmotions dataset, Logistic Regression and Linear Support Vector Machines performed reliably on high-frequency emotions such as neutral, joy, and approval. However, per-class evaluation revealed consistent degradation in performance for low resource emotions including grief, relief, and remorse. Error analysis showed that these failures were primarily due to severe label imbalance and semantic similarity between emotions, rather than random model errors.

Results on the WASSA dataset highlighted the impact of short, informal social media text on emotion classification. While hyper parameter optimized Logistic Regression and LinearSVC models significantly outperformed the baseline, confusion matrices revealed systematic misclassification between sadness and fear. Manual inspection of misclassified examples indicated that sarcasm, ambiguous phrasing, and emoji usage were dominant sources of error.

Experiments on the Lexicon dataset confirmed the limitations of single-word emotion classification. Despite improvements achieved with classifiers, performance was constrained with high overlapping issues between emotion categories. Low-frequency emotions such as surprise and anticipation were particularly difficult to model.

Comparison of Some of the Models and Datasets Based on Project Findings

Dataset	Model / Approach	What the Model Performed Well On	Observed Weaknesses & Failure Modes	Evidence from Error Analysis
GoEmotions	TF-IDF & Logistic Regression (Baseline)	Correctly identifies frequent emotions such as neutral, joy, and approval	Struggles with rare emotions (grief, relief, remorse); poor handling of mixed emotions	Lowest per-class F1 scores concentrated on rare labels; high false negatives for low-support classes
GoEmotions	TF-IDF & LinearSVC (HPO)	Improved classification of excitement and confusion compared to baseline and LR	Still misclassifies subtle or overlapping emotions (pride vs. approval, annoyance vs. anger)	Label-pair confusion analysis shows frequent swaps among semantically close emotions
WASSA	TF-IDF & DummyClassifier (Baseline)	Establishes majority-class reference performance	Fails almost entirely to distinguish emotional nuance	Macro-F1 near zero, confirming need for learned models
WASSA	TF-IDF + Logistic Regression (HPO)	Strong improvement over baseline; captures explicit emotion cues in tweets	Confuses *sadness* and fear when intensity cues are weak	Confusion matrix shows dominant off-diagonal errors between sadness and fear
WASSA	TF-IDF & LinearSVC (HPO)	Best overall F1 performance among classical models	Misclassifies sarcastic tweets and emoji-driven expressions	Manual inspection highlights sarcasm and irony as dominant error sources
Lexicon	Majority Class Baseline	Establishes frequency-based reference performance	Predicts dominant emotions only; zero recall for minority emotions	Classification report shows near-zero F1 for most emotions
Lexicon	CountVectorizer + Logistic Regression	Captures character-level emotional cues	Over predicts common emotions	High false positives for negative and positive labels
Lexicon	TF-IDF + Logistic Regression	Improved balance across frequent emotions	Still struggles with overlapping emotion labels	FP/FN analysis shows confusion among fear, anger, negative
Lexicon	TF-IDF + LinearSVC (HPO)	Best overall performance among lexicon models	Can not resolve ambiguity without context	Common misclassified words shared across all models

Future work could explore adaptive decision thresholds for multi-label classification to address the prediction of rare emotions. Incorporating emotion-specific weighting schemes or data augmentation techniques may help solve class imbalance issues. Additionally, fine-tuning transformer models could improve performance on subtle emotional expressions. Finally, extending the analysis to cross-dataset generalization would provide better insight into model across different social media domains.

5. Conclusion

This project evaluated multiple Natural Language Processing approaches for emotion detection in social media text using the GoEmotions, WASSA, and NRC Lexicon datasets. The results show that baseline models are poor but perform well on frequent emotions, while HPO models provide improved contextual understanding but still struggle with rare and overlapping emotion categories.

Detailed error analysis revealed that class imbalance, semantic similarity between emotions, and informal language are the primary challenges in emotion classification. Overall, this work highlights both the effectiveness and limitations of current emotion detection methods and emphasizes the importance of context aware modeling and careful evaluation.

6. References

- Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). *GoEmotions: A Dataset of Fine-Grained Emotions*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).
- Mohammad, S. M., & Turney, P. D. (2013). *Crowdsourcing a Word–Emotion Association Lexicon*. Computational Intelligence, 29(3), 436–465.
- Abdul-Mageed, M., & Ungar, L. (2017). *EmoInt at EmoInt-2017: Emotion Intensity Prediction Using Affective Lexicons and Deep Learning*. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA).