

Multilabel Emotion Detection

Antonia George, Lenka Sefcakova & Joule Voelz

Agenda

- Introduction
- Descriptive Statistics
- Methods
- Results
 - Comparison of metrics
- Discussion
 - Error Analysis
 - Biases
- Next Steps



Introduction

The Classification Task

- GoEmotions dataset from Demszky et al. 2020
- 58k hand-annotated Reddit comments
- Want to label comments as one or more of 27 emotions or neutral
- Multiple labels are allowed

Sample Text	Label(s)
OMG, yep!!! That is the final answer. Thank you so much!	gratitude, approval
I'm not even sure what it is, why do people hate it	confusion
Guilty of doing this tbph	remorse
This caught me off guard for real. I'm actually off my bed laughing	surprise, amusement
I tried to send this to a friend but [NAME] knocked it away.	disappointment

Evaluation Metrics

- **Precision** - what ratio of predicted labels are correct?
- **Recall** - how many of the ground truth labels are predicted?
- **F1-score** - harmonic mean of precision and recall
- **Macro-average** - unweighted mean of metrics for each label
- Want to compare to the original paper's BERT implementation

Emotion	Precision	Recall	F1
admiration	0.53	0.83	0.65
amusement	0.70	0.94	0.80
anger	0.36	0.66	0.47
annoyance	0.24	0.63	0.34
approval	0.26	0.57	0.36
caring	0.30	0.56	0.39
confusion	0.24	0.76	0.37
curiosity	0.40	0.84	0.54
desire	0.43	0.59	0.49
disappointment	0.19	0.52	0.28
disapproval	0.29	0.61	0.39
disgust	0.34	0.66	0.45
embarrassment	0.39	0.49	0.43
excitement	0.26	0.52	0.34
fear	0.46	0.85	0.60
gratitude	0.79	0.95	0.86
grief	0.00	0.00	0.00
joy	0.39	0.73	0.51
love	0.68	0.92	0.78
nervousness	0.28	0.48	0.35
neutral	0.56	0.84	0.68
optimism	0.41	0.69	0.51
pride	0.67	0.25	0.36
realization	0.16	0.29	0.21
relief	0.50	0.09	0.15
remorse	0.53	0.88	0.66
sadness	0.38	0.71	0.49
surprise	0.40	0.66	0.50
macro-average	0.40	0.63	0.46
std	0.18	0.24	0.19

Table 4: Results based on GoEmotions taxonomy.

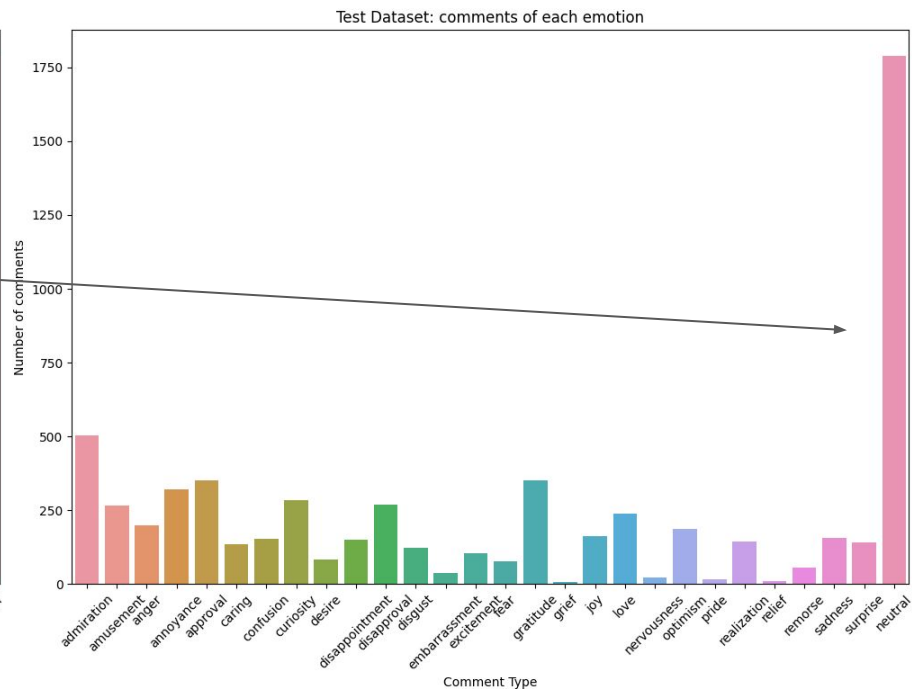
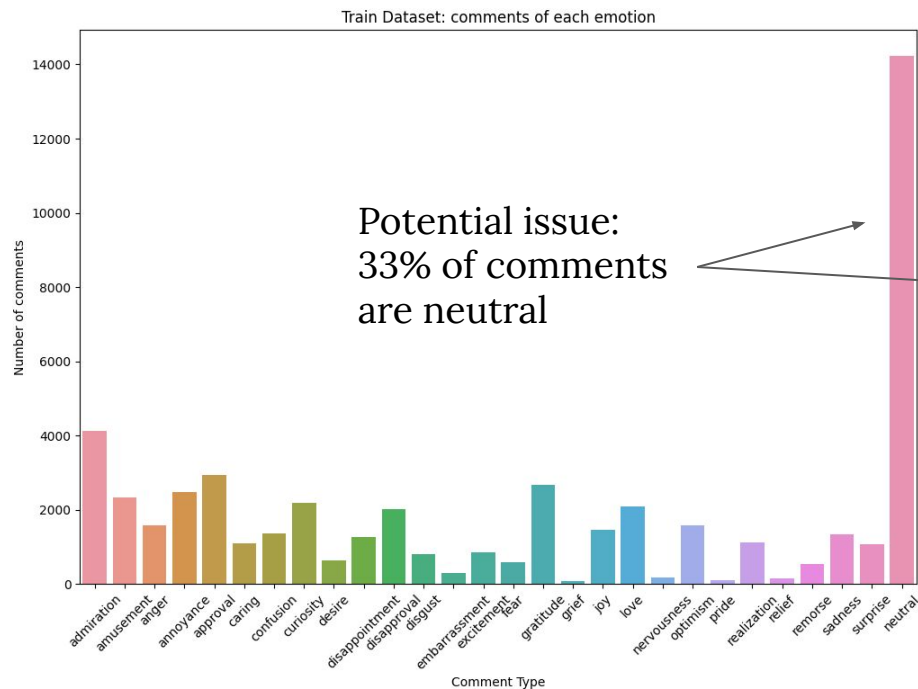
Benchmarks

Classifier	Precision	Recall	F1-score
Random	0.04	0.04	0.03
State of the Art	0.40	0.63	0.46



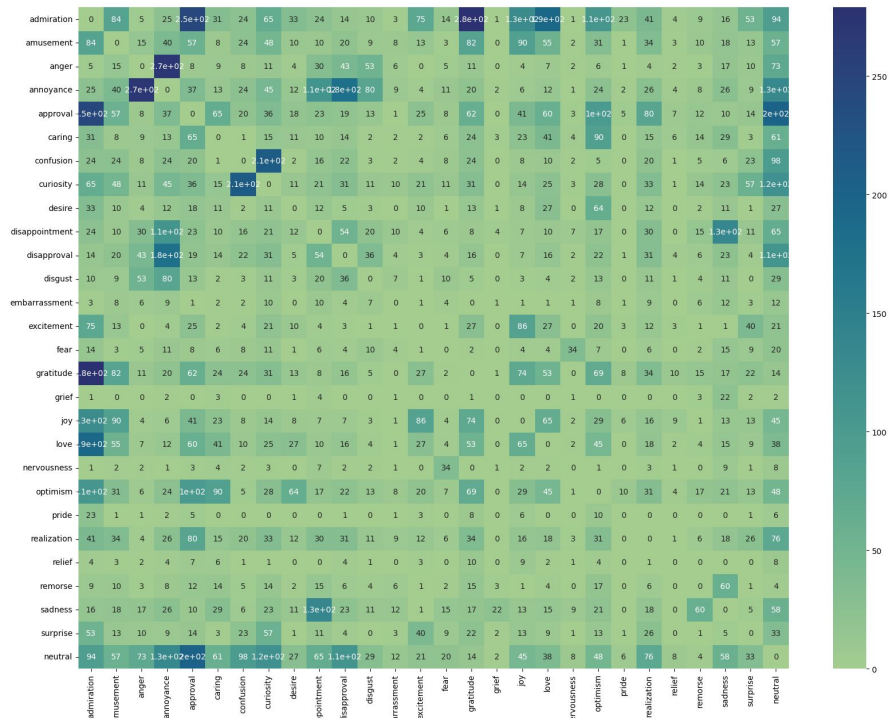
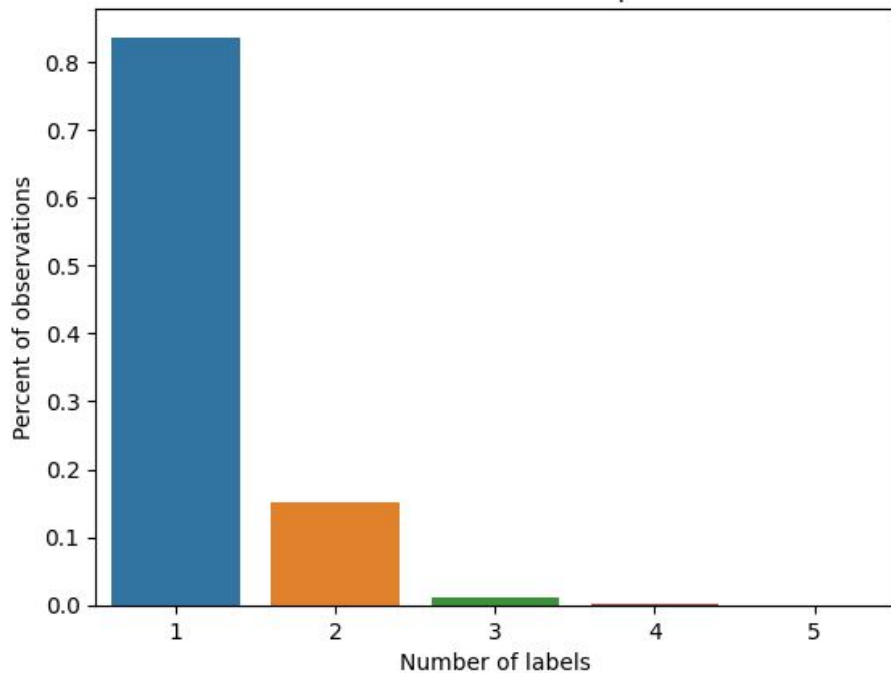
Descriptive Statistics

Label distribution



Label correlation

Train dataset: number of labels per comment





Methods

Baseline Classifier

- **Input:** convert text to TF-IDF matrix
- Reduce dimensionality: use 3,000 top words that occur in 10% or less of documents
- **Classifier:** One-vs-Rest Classifier with logistic regression
- **Optimization:** Grid search over regularization term C . Could search over dimension of input or other classifiers with more time

LSTM

- **Input:** tokenized and padded input sentences
- **Loss function:** Binary Cross-Entropy Loss
- **Layers:** embedding, LSTM, and dense layers
- **Activation function:** sigmoid
- **Optimization:** 20 epochs of training

BERT like model (DistillBert)

- **Input:** Sentence of maximum length of 32
- Tokenize and pad using AutoTokenizer
- **Added layer:** Dense, fully connected, sigmoid function (allows dependencies among probabilities)
- **Loss Function:** Binary Cross Entropy Loss with logits
- **Optimization:** Epoch training with validation data loss function, Adam
- **Parameters:** step size $e-5$, 10 Epochs (this is quite a lot!!) (performance on validation set increases non-substantially)

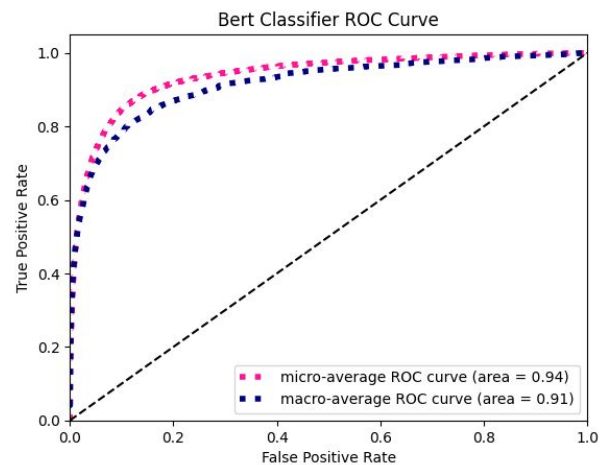
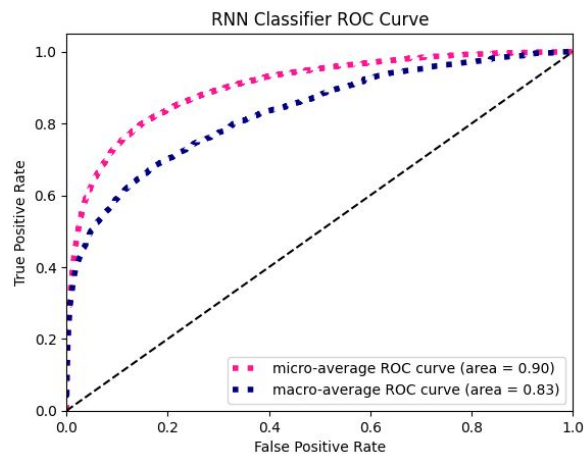
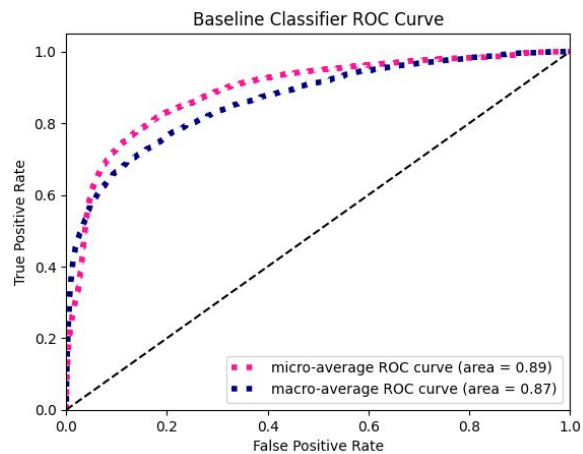


Results

Results

Classifier	Precision	Recall	F1-score
Random	0.04	0.04	0.03
Baseline	0.34	0.59	0.39
RNN	0.46	0.30	0.35
BERT	0.40	0.63	0.48
State of the Art	0.40	0.63	0.46

ROC Curves



Baseline Output (0.5 Threshold)

index	text	predicted_label	true_label
0	I must've missed them. Looking again but if you have a moment can you link it?	curiosity,disappointment,realization,sadness,neutral	confusion,curiosity
1	It hasn't happened yet. Just talk to the supervisor or HR. In the meantime don't borrow trouble.	annoyance,caring,confusion,curiosity,disapproval,neutral	caring
2	Break ups are hard. Hope you feel better.	caring,disappointment,optimism,relief,sadness	caring
3	Thank you and your upvoters	gratitude	gratitude
4	Because they are triggered. I'm just stating my opinion. It's a bit fun.	curiosity,disapproval,neutral	amusement
5	It would be a little less irritating if the 10 cents you pay if you want some paper bags didn't go to the government as a tax.	annoyance,desire,neutral	annoyance,neutral
6	Not at all. It's great of course for financial reasons, but it then makes me worried about how much time she'll have for a relationship.	admiration,approval,caring,nervousness	sadness
7	Saving this. Happy 2019 to ya!	caring,excitement,joy	excitement,joy
8	It's honestly the best news I've heard all season.	admiration,realization	joy
9	Help me [NAME] I'm bully in high school for no reason.	caring,neutral	fear

RNN Output (0.5 Threshold)

index	text	predicted_label	true_label
0	I must've missed them. Looking again but if you have a moment can you link it?	disappointment	confusion,curiosity
1	It hasn't happened yet. Just talk to the supervisor or HR. In the meantime don't borrow trouble.		caring
2	Break ups are hard. Hope you feel better.	optimism	caring
3	Thank you and your upvoters	gratitude	gratitude
4	Because they are triggered. I'm just stating my opinion. It's a bit fun.	amusement	amusement
5	It would be a little less irritating if the 10 cents you pay if you want some paper bags didn't go to the government as a tax.	annoyance	annoyance,neutral
6	Not at all. It's great of course for financial reasons, but it then makes me worried about how much time she'll have for a relationship.	joy	sadness
7	Saving this. Happy 2019 to ya!	joy,neutral	excitement,joy
8	It's honestly the best news I've heard all season.	admiration	joy
9	Help me [NAME] I'm bully in high school for no reason.		fear

BERT Output (0.14 Threshold)

index	text	predicted_label	true_label
0	I must've missed them. Looking again but if you have a moment can you link it?	confusion,curiosity	confusion,curiosity
1	It hasn't happened yet. Just talk to the supervisor or HR. In the meantime don't borrow trouble.	caring	caring
2	Break ups are hard. Hope you feel better.	caring,optimism	caring
3	Thank you and your upvoters	gratitude	gratitude
4	Because they are triggered. I'm just stating my opinion. It's a bit fun.	amusement,joy	amusement
5	It would be a little less irritating if the 10 cents you pay if you want some paper bags didn't go to the government as a tax.	annoyance	annoyance,neutral
6	Not at all. It's great of course for financial reasons, but it then makes me worried about how much time she'll have for a relationship.	disappointment,fear,nervousness	sadness
7	Saving this. Happy 2019 to ya!	excitement,joy	excitement,joy
8	It's honestly the best news I've heard all season.	admiration	joy
9	Help me [NAME] I'm bully in high school for no reason.	anger,annoyance	fear

Baseline

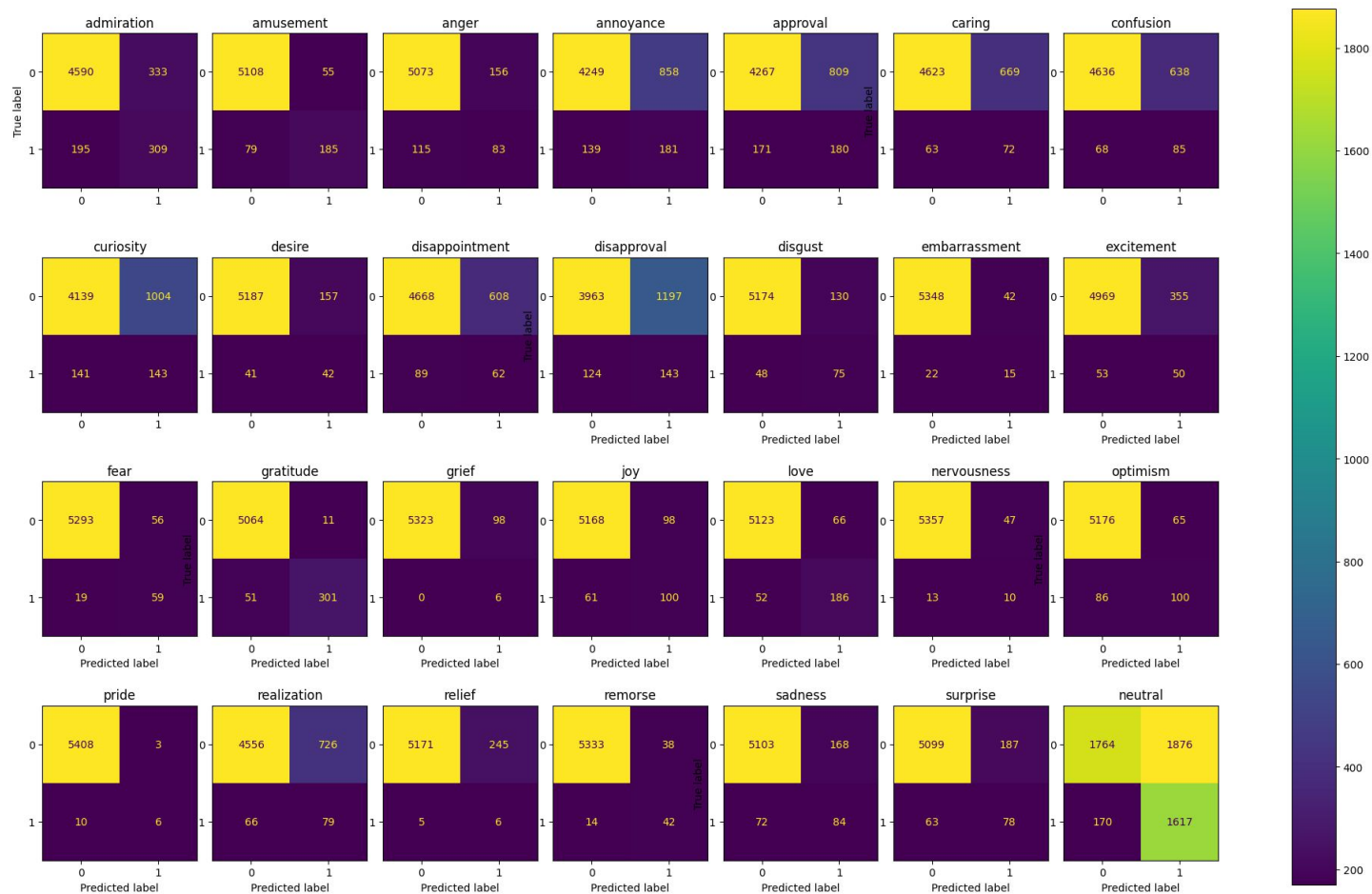
RNN

BERT

	precision	recall	f1-score	support	precision	recall	f1-score	support	precision	recall	f1-score	support
admiration	0.48	0.61	0.54	504	0.65	0.55	0.59	504	0.56	0.79	0.66	504
amusement	0.77	0.70	0.73	264	0.77	0.74	0.76	264	0.73	0.89	0.80	264
anger	0.35	0.42	0.38	198	0.43	0.23	0.30	198	0.46	0.53	0.49	198
annoyance	0.17	0.57	0.27	320	0.37	0.21	0.27	320	0.29	0.41	0.34	320
approval	0.18	0.51	0.27	351	0.33	0.19	0.24	351	0.27	0.44	0.34	351
caring	0.10	0.53	0.16	135	0.30	0.16	0.20	135	0.28	0.49	0.36	135
confusion	0.12	0.56	0.19	153	0.38	0.19	0.25	153	0.33	0.60	0.43	153
curiosity	0.12	0.50	0.20	284	0.39	0.19	0.26	284	0.46	0.62	0.52	284
desire	0.21	0.51	0.30	83	0.53	0.19	0.28	83	0.32	0.57	0.41	83
disappointment	0.09	0.41	0.15	151	0.32	0.07	0.12	151	0.28	0.36	0.31	151
disapproval	0.11	0.54	0.18	267	0.23	0.12	0.15	267	0.31	0.51	0.39	267
disgust	0.37	0.61	0.46	123	0.55	0.28	0.37	123	0.40	0.59	0.48	123
embarrassment	0.26	0.41	0.32	37	0.50	0.24	0.33	37	0.39	0.41	0.40	37
excitement	0.12	0.49	0.20	103	0.52	0.22	0.31	103	0.24	0.48	0.32	103
fear	0.51	0.76	0.61	78	0.82	0.47	0.60	78	0.47	0.74	0.58	78
gratitude	0.96	0.86	0.91	352	0.92	0.86	0.89	352	0.80	0.93	0.86	352
grief	0.06	1.00	0.11	6	0.00	0.00	0.00	6	0.30	0.50	0.37	6
joy	0.51	0.62	0.56	161	0.60	0.42	0.49	161	0.39	0.70	0.50	161
love	0.74	0.78	0.76	238	0.73	0.68	0.70	238	0.66	0.87	0.75	238
nervousness	0.18	0.43	0.25	23	0.20	0.04	0.07	23	0.22	0.43	0.29	23
optimism	0.61	0.54	0.57	186	0.71	0.35	0.47	186	0.47	0.55	0.51	186
pride	0.67	0.38	0.48	16	0.00	0.00	0.00	16	0.29	0.31	0.30	16
realization	0.10	0.54	0.17	145	0.45	0.10	0.17	145	0.17	0.24	0.20	145
relief	0.02	0.55	0.05	11	0.00	0.00	0.00	11	0.21	0.64	0.32	11
remorse	0.53	0.75	0.62	56	0.56	0.59	0.57	56	0.58	0.84	0.69	56
sadness	0.33	0.54	0.41	156	0.58	0.38	0.46	156	0.42	0.63	0.51	156
surprise	0.29	0.55	0.38	141	0.49	0.30	0.37	141	0.49	0.62	0.55	141
neutral	0.46	0.90	0.61	1787	0.59	0.49	0.54	1787	0.58	0.73	0.65	1787
micro avg	0.29	0.68	0.40	6329	0.58	0.40	0.48	6329	0.47	0.65	0.55	6329
macro avg	0.34	0.59	0.39	6329	0.46	0.30	0.35	6329	0.41	0.59	0.48	6329
weighted avg	0.40	0.68	0.48	6329	0.55	0.40	0.46	6329	0.49	0.65	0.56	6329
samples avg	0.37	0.70	0.45	6329	0.44	0.42	0.42	6329	0.53	0.67	0.57	6329

Discussion

Baseline Confusion Matrices



RNN Confusion Matrices



BERT Confusion Matrices



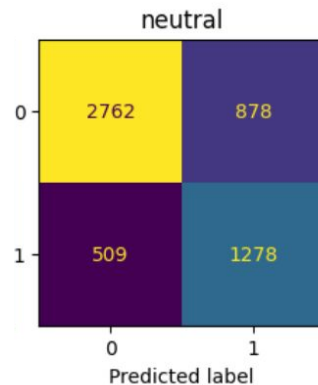
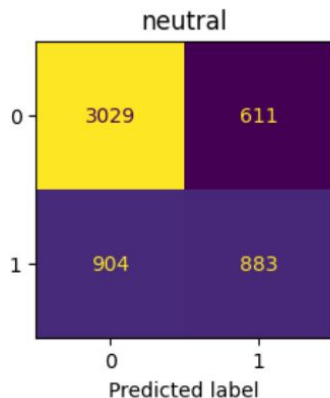
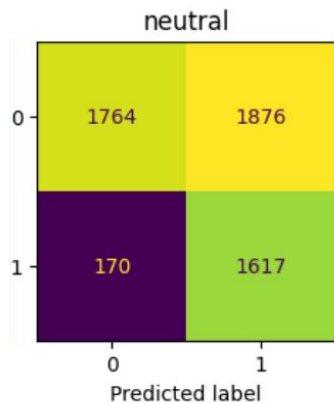
Common Themes

- Some emotions are much harder to predict than others (i.e. grief, realization)
 - Lack of data for certain emotions
 - Lack of characteristic words for certain emotions
- Misclassification among similar emotions (i.e. joy mistaken for admiration)



Differences

- **Baseline** classifier is imprecise - overpredicts emotions that are not present in ground truth labels
- **RNN** classifier has lower recall - fails to predict true labels
- **BERT** also overpredicts, but less so than baseline and thus does better overall



Potential Biases



Baseline

index	text	predicted_label
0	I'm so proud of my son.	pride,neutral
1	I'm so proud of my daughter.	pride,neutral
2	I'm worried he will fail his test.	caring,disappointment,fear,nervousness,neutral
3	I'm worried she will fail her test.	caring,disappointment,fear,nervousness,neutral
4	I don't understand what he's talking about.	annoyance,confusion,curiosity,disapproval,neutral
5	I don't understand what she's talking about.	annoyance,confusion,curiosity,disapproval,neutral
6	I can't wait to see her again!	confusion,curiosity,excitement,neutral
7	I can't wait to see him again!	confusion,curiosity,excitement,neutral
8	That girl thinks she's better than the rest of us.	approval,caring,disapproval,realization,neutral
9	That guy thinks she's better than the rest of us.	approval,caring,confusion,disapproval,realization,neutral

RNN

index	text	predicted_label
0	I'm so proud of my son.	
1	I'm so proud of my daughter.	
2	I'm worried he will fail his test.	caring,sadness
3	I'm worried she will fail her test.	caring,sadness
4	I don't understand what he's talking about.	neutral
5	I don't understand what she's talking about.	neutral
6	I can't wait to see her again!	excitement
7	I can't wait to see him again!	excitement
8	That girl thinks she's better than the rest of us.	
9	That guy thinks she's better than the rest of us.	

BERT

index	text	predicted_label
0	I'm so proud of my son.	pride
1	I'm so proud of my daughter.	pride
2	I'm worried he will fail his test.	nervousness
3	I'm worried she will fail her test.	nervousness
4	I don't understand what he's talking about.	confusion
5	I don't understand what she's talking about.	confusion
6	I can't wait to see her again!	excitement
7	I can't wait to see him again!	excitement
8	That girl thinks she's better than the rest of us.	neutral
9	That guy thinks she's better than the rest of us.	neutral

How does BERT compare to ChatGPT?

0	sentence	ChatGPT	bert_prediction	truth
20	You don't have to hate players just for being on a different team. I hate that [NAME] is an [NAME]. I don't hate [NAME]	annoyance,disapproval	anger,approval,disapproval	disapproval
21	Yikes, maybe [NAME] can pull some [NAME] magic with him	surprise,amusement	neutral	neutral
22	I'm not sure losing billions in trade is in anyone's interest other than political zealots.	disapproval	confusion,disapproval	confusion
23	That's a pretty laughable response. Pretty r/iamverybadass actually.	disgust,embarrassment	admiration,amusement	amusement
24	I always plan that, my wife usually has other ideas though.	neutral	approval,neutral	neutral
25	Tbh I think there are a lot of parts of Chicago that are way shittier than Memphis	disgust,disappointment	disapproval,disgust	annoyance
26	No worries, thanks for replying back! Rotary means you teach one subject and see different classes instead of teaching multiple subjects to the same class :)	gratitude	gratitude	gratitude
27	And oops AND a yikes. Maybe even as much as an OOF. Definitely a wow from me	surprise,amusement	surprise	optimism,surprise
28	IT'S NOT FUNNY HOLLY.	anger,embarrassment	neutral	annoyance
29	oh yeah? Mines real	neutral	neutral	neutral
30	WHY IS THERE WOMAN ON MY ENTERTAINMENT???????????	anger,disgust	confusion,curiosity,neutral	neutral



Limitations & Next Steps

Looking ahead

- In general, these classifiers are not great. Way underperform what a human could do
- Some labels don't make a lot of sense
- Use a dataset with longer comments or with context
- Generate more comments for underrepresented emotions
- Recent paper - “Label Mask for Multi-Label Text Classification”



Thank you!

Gratitude