Capstone 3 – Emotion Project Summary Report

This project was conducted to answer the problem question:

Can you create a model to do ‘emotion labeling’ accurately and consistently given relevant text?

Context

‘Emotion labeling’ is a powerful tool in crisis intervention, psychotherapy. By labeling the specific emotion that an individual is expressing, the individual can often regulate their feelings and gain a deeper understanding of their emotional situation.

Data Wrangling

Google Research Team had created a dataset (GoEmotions) of 58,000 carefully curated comments extracted from Reddit, with human annotations to 27 emotion categories or Neutral. There was an additional version that was compiled using rater-agreement to create the dataset used in this project. This was already broken into a training and testing split which was maintained to avoid any data leakage.

GoEmotions Database: https://github.com/google-research/google-research/blob/master/goemotions/README.md

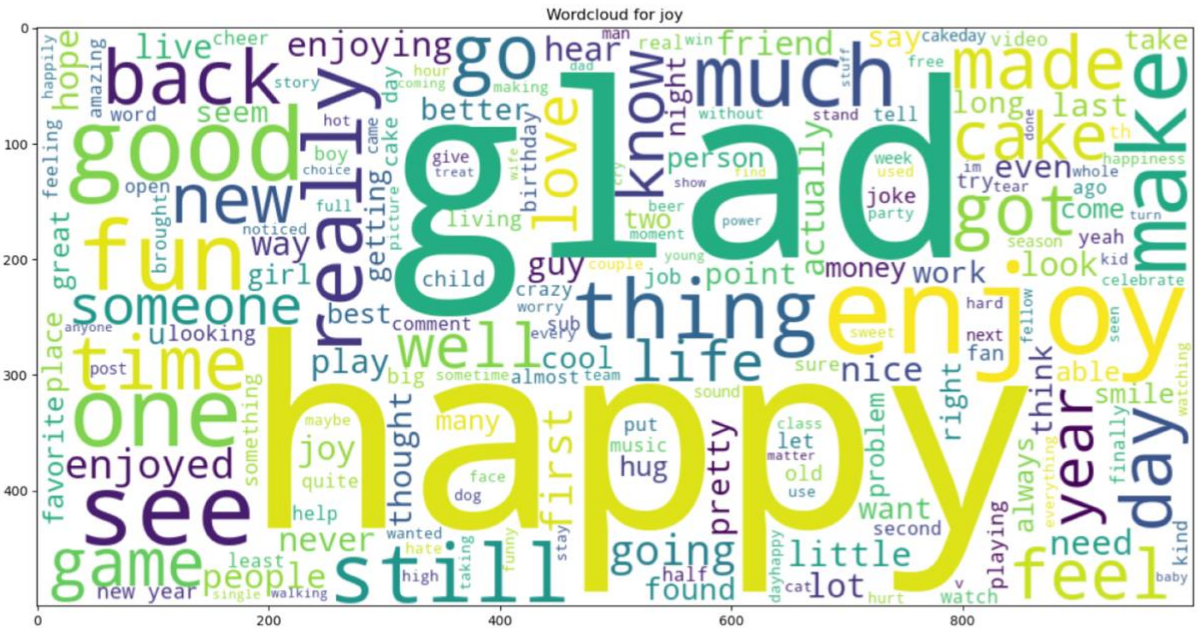
This dataset had three columns: original text, list of emotion labels assigned to this text by raters, and id of the poster. The original list of emotions (numbers) was mapped back to see the corresponding emotions.

Natural language processing techniques were used to break down the original text into more uniform text. The text was then lemmatized for further standardization. Upon further exploration, it was seen that any identifying nouns in the original text were replaced with ‘[NAME]’. So, ‘name’ was also added to the list of stop words because this artificially caused ‘name’ to become very prominent in the original text.

Exploratory Data Analysis

The NLP-cleaned text was then used to create word clouds for each emotion. This provided a visual sense of the most common words in each emotional label. Below are two example word clouds for the emotions joy and grief.

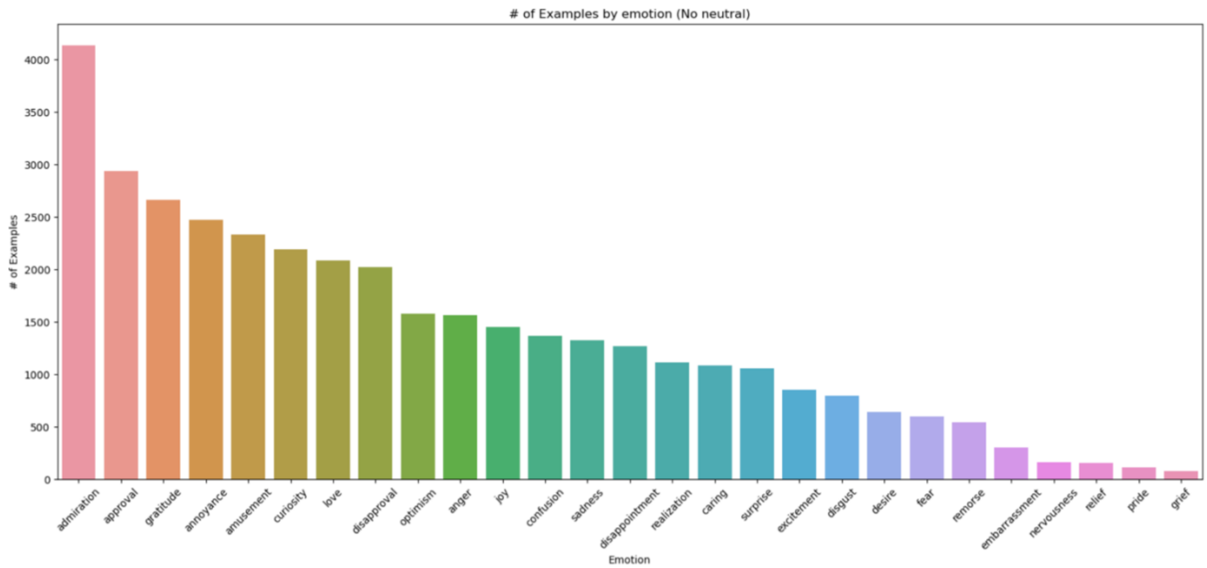
Word cloud for Joy



Word cloud for Grief

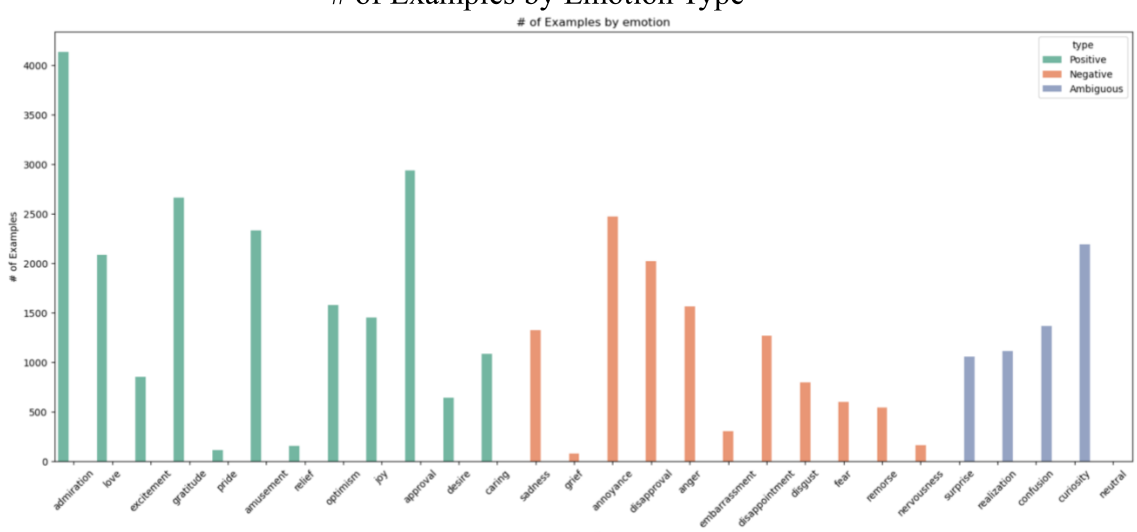
The following visual shows the breakdown of the dataset by number of examples of each emotion. Neutral emotions were removed because they were by far the most represented example. The chart shows moderate spread in examples with some emotions (grief, pride) having much fewer examples. This imbalance will likely affect how a model trains and predicts on text.

# of Examples by Emotion (No neutral)

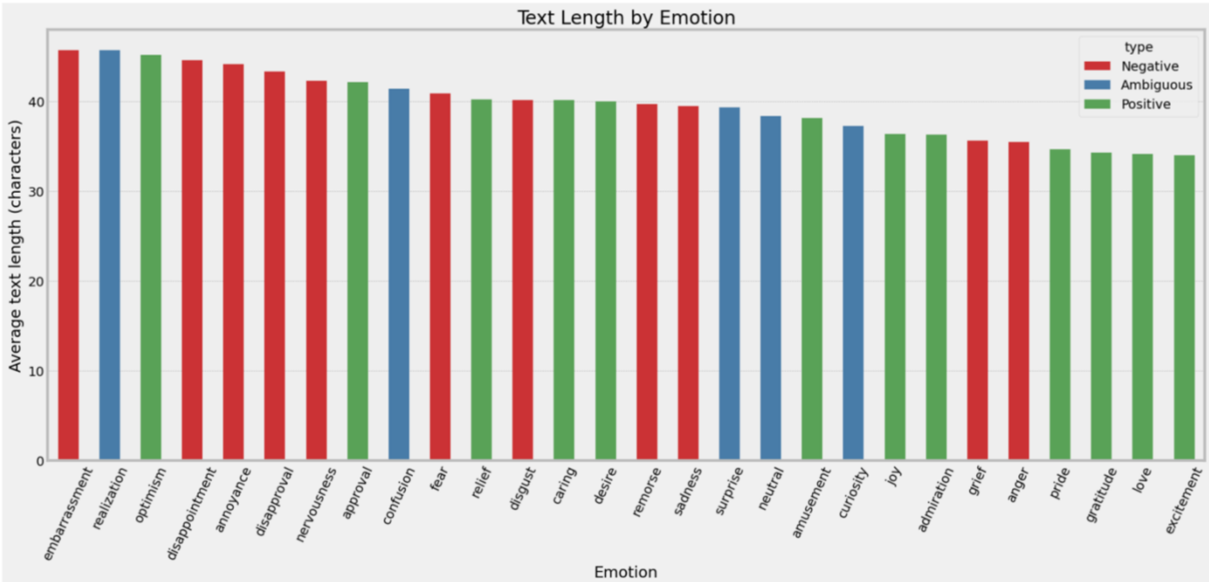


An additional grouping was made assigning each emotion as either ‘positive’, ‘negative’, or ‘ambiguous’. The following graph shows there are more positive emotions and more examples of them also compared to negative and ambiguous emotions.

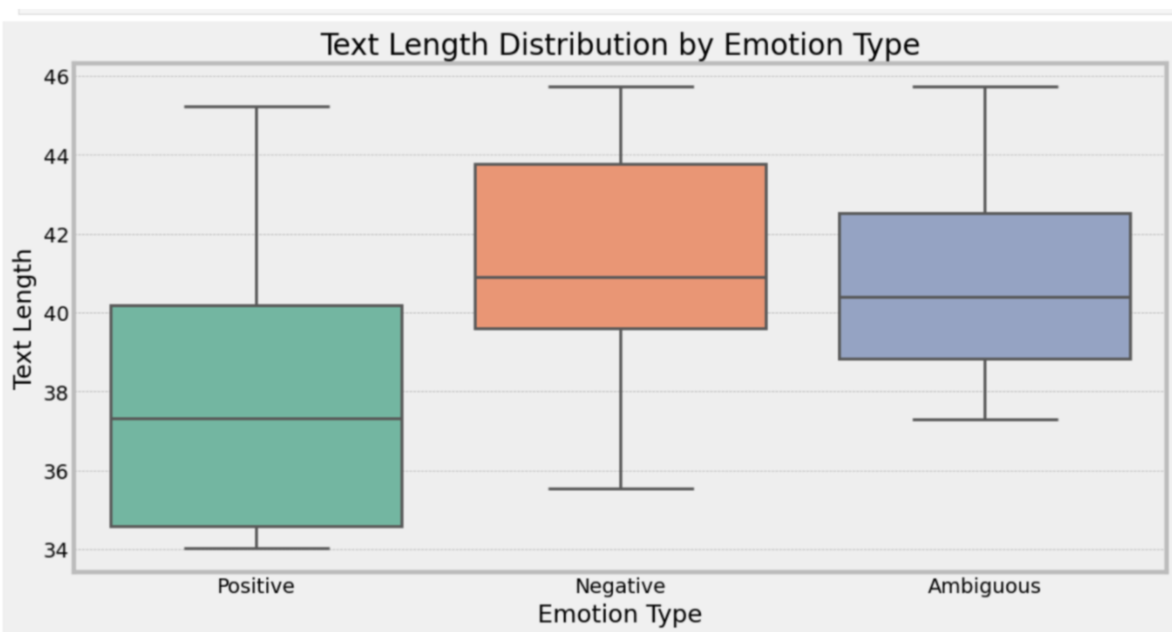
# of Examples by Emotion Type



The length of each text was also analyzed for possible patterns. There appears to be a slight trend of negative emotion texts being longer in length.



The following box plot provides a closer look at the distribution of text length by category of emotion. Negative emotion texts do in fact appear to be longer than positive emotion texts.



Modeling

The main goal of this project was to build a model that can accurately identify emotions given new text if applicable.Prior to the modeling step, some pre-processing was done. One-hot encoding was done to create new binary columns for each emotion.

Undersampling/Oversampling considerations

One of the major considerations was whether to use oversampling/undersampling because of the imbalanced number of examples per emotion in the dataset. However, the dataset was large (58,000 texts) and samples from online at random. Also, in the context of human conversation and even more specifically in mental health settings, all emotions are not equally represented. It is realistic that certain emotions will appear much more than others and so it is fitting that this model should also have a similar representation. As a result, over and undersampling emotions were not used to train the models.

Scoring Metric

The other key decision was to decide the scoring metric to compare different emotions. The purpose of this model is to be used in a chatbot that can read new text and label relevant emotions if appropriate. In this context, it is important that when the model labels an emotion, it is the correct emotion. For example, a therapist listening to a sad statement from a client and responding as if it was a joyous statement would be highly inappropriate. On the flip side, it is less imperative that the model recognizes every single instance of an emotion because in this context, it is likely that successive texts will often be of the same emotion. For example, even if a therapist does not recognize a client's statement as sadness, often the client's consequent statements will also be of sadness.
As a result, micro precision was the key metric to be prioritized with micro recall not prioritized as much. Ideally, the model will also recognize most if not all models present in this dataset.

To prepare the dataset for modeling, the NLP-cleaned text was then broken down using Count Vectorizer. Tfidf-Vectorizer was also briefly explored but it gave worse results during the modeling stage.

Multiple model types were trained and tested and evaluated. The following graphics show classification reports for naïve Bayes model, random forest model, logistic regression model and K-nearest neighbors model.

Naïve Bayes Model

```
Evaluating the model...
This is the classification report for {'classifier': OneVsRestClassifier(estimator=MultinomialNB())}
```

	precision	recall	f1-score	support
admiration	0.85	0.08	0.15	504
amusement	1.00	0.03	0.07	264
anger	0.00	0.00	nan	198
annoyance	1.00	0.00	0.01	320
approval	0.00	0.00	nan	351
caring	nan	0.00	nan	135
confusion	0.00	0.00	nan	153
curiosity	nan	0.00	nan	284
desire	nan	0.00	nan	83
disappointment	nan	0.00	nan	151
disapproval	nan	0.00	nan	267
disgust	1.00	0.02	0.03	123
embarrassment	nan	0.00	nan	37
excitement	1.00	0.02	0.04	103
fear	nan	0.00	nan	78
gratitude	0.98	0.39	0.56	352
grief	nan	0.00	nan	6
joy	0.25	0.01	0.01	161
love	1.00	0.06	0.12	238
nervousness	nan	0.00	nan	23
optimism	0.50	0.01	0.01	186
pride	nan	0.00	nan	16
realization	nan	0.00	nan	145
relief	nan	0.00	nan	11
remorse	nan	0.00	nan	56
sadness	1.00	0.01	0.01	156
surprise	1.00	0.01	0.01	141
neutral	0.61	0.11	0.19	1787
micro avg	0.74	0.06	0.12	6329
macro avg	0.68	0.03	0.10	6329
weighted avg	0.66	0.06	0.15	6329
samples avg	0.74	0.07	0.95	6329

Random Forest

```
This is the classification report for {'classifier': OneVsRestClassifier(estimator=RandomForestClassifier(max
_samples=20000,
                                                                   n_estimators=50))}
```

	precision	recall	f1-score	support
admiration	0.66	0.53	0.59	504
amusement	0.81	0.64	0.72	264
anger	0.50	0.25	0.33	198
annoyance	0.71	0.09	0.16	320
approval	0.56	0.08	0.14	351
caring	0.65	0.10	0.17	135
confusion	0.59	0.07	0.12	153
curiosity	0.67	0.04	0.07	284
desire	0.83	0.18	0.30	83
disappointment	0.50	0.01	0.03	151
disapproval	0.29	0.01	0.03	267
disgust	0.93	0.20	0.33	123
embarrassment	nan	0.00	nan	37
excitement	0.43	0.03	0.05	103
fear	0.73	0.14	0.24	78
gratitude	0.95	0.87	0.91	352
grief	nan	0.00	nan	6
joy	0.67	0.22	0.33	161
love	0.76	0.73	0.74	238
nervousness	nan	0.00	nan	23
optimism	0.73	0.40	0.52	186
pride	0.50	0.06	0.11	16
realization	0.33	0.01	0.01	145
relief	nan	0.00	nan	11
remorse	0.65	0.20	0.30	56
sadness	0.69	0.29	0.41	156
surprise	0.58	0.10	0.17	141
neutral	0.59	0.55	0.57	1787
micro avg	0.67	0.36	0.47	6329
macro avg	0.64	0.21	0.31	6329
weighted avg	0.63	0.36	0.41	6329
samples avg	0.67	0.38	0.94	6329

Logistic Regression

	precision	recall	f1-score	support
admiration	0.71	0.45	0.55	504
amusement	0.80	0.71	0.75	264
anger	0.48	0.13	0.20	198
annoyance	0.55	0.07	0.12	320
approval	0.49	0.11	0.18	351
caring	0.42	0.08	0.14	135
confusion	0.40	0.05	0.09	153
curiosity	0.62	0.04	0.07	284
desire	0.65	0.20	0.31	83
disappointment	0.80	0.05	0.10	151
disapproval	0.50	0.02	0.04	267
disgust	0.72	0.24	0.36	123
embarrassment	0.57	0.11	0.18	37
excitement	0.71	0.19	0.31	103
fear	0.79	0.42	0.55	78
gratitude	0.95	0.88	0.91	352
grief	1.00	0.17	0.29	6
joy	0.60	0.48	0.54	161
love	0.76	0.69	0.72	238
nervousness	0.00	0.00	0.00	23
optimism	0.74	0.46	0.57	186
pride	1.00	0.06	0.12	16
realization	0.62	0.07	0.12	145
relief	0.00	0.00	0.00	11
remorse	0.53	0.55	0.54	56
sadness	0.66	0.29	0.40	156
surprise	0.47	0.29	0.36	141
neutral	0.46	0.92	0.61	1787
micro avg	0.55	0.48	0.51	6329
macro avg	0.61	0.28	0.33	6329
weighted avg	0.59	0.48	0.44	6329
samples avg	0.55	0.51	0.52	6329

K- Nearest Neighbors

```
This is the classification report for {'classifier': OneVsRestClassifier(estimator=KNeighborsClassifier(n_nei
ghbors=7, p=1,
                                                                   weights='distance'))}
```

	precision	recall	f1-score	support
admiration	0.62	0.34	0.44	504
amusement	0.82	0.36	0.50	264
anger	0.47	0.18	0.26	198
annoyance	0.29	0.02	0.03	320
approval	0.35	0.05	0.09	351
caring	0.50	0.07	0.13	135
confusion	0.06	0.01	0.01	153
curiosity	0.00	0.00	nan	284
desire	0.67	0.12	0.20	83
disappointment	0.00	0.00	nan	151
disapproval	0.00	0.00	nan	267
disgust	0.94	0.13	0.23	123
embarrassment	0.00	0.00	nan	37
excitement	0.29	0.02	0.04	103
fear	0.88	0.09	0.16	78
gratitude	0.95	0.80	0.87	352
grief	nan	0.00	nan	6
joy	0.60	0.22	0.32	161
love	0.54	0.71	0.61	238
nervousness	nan	0.00	nan	23
optimism	0.65	0.15	0.24	186
pride	0.50	0.06	0.11	16
[illegible]	[illegible]	0.00	nan	145
relief	nan	0.00	nan	11
remorse	0.65	0.39	0.49	56
sadness	0.72	0.15	0.24	156
surprise	0.42	0.06	0.10	141
neutral	0.49	0.64	0.55	1787
micro avg	0.55	0.33	0.41	6329
macro avg	0.46	0.16	0.28	6329
weighted avg	0.47	0.33	0.40	6329
samples avg	0.55	0.35	0.94	6329

Summary statistics for models

	micro_precision	micro_recall
Naive Bayes	0.74	0.06
Random Forest	0.67	0.36
Logistic Regression	0.55	0.48
KNN	0.55	0.33

Looking at the different models here, Random Forest provides the best precision while still identifying most of the emotions. While Naive Bayes has a higher precision in theory, it does so by simply ignoring most of the emotions (shown by the very low recall number). The advantage of random forest is that it can also continue to be adapted for more emotions with appropriately scaling complexity.

Future Work/Additional Info/Considerations

The next step in this project is to design a chatbot that can incorporate this model as part of larger framework in a crisis text/mental health setting. Specific to emotion labeling, more datasets representing other emotions, other sources of texts and ideally texts from a crisis setting could all help improve this model.

One of the notable limitations of this project is that the texts are sourced from Reddit which is not necessarily representative of the larger population. Also, most of this text is likely dissimilar from text that would be present in a crisis response/mental health setting. Also, there are significantly more emotions that could be labeled and used to provide more robust support.