

Mini-Project 1

4. Analysis

4.1. The provided dataset has certain particularities that will have an impact on the performance of our models. First and foremost, the dataset is extremely unbalanced, with the emotion 'neutral' making over 50% of the entire dataset. Furthermore, many of the words are misspelled and we find a lot of random words that have been tokenized such as number sequences and kanji and mandarin characters. Although their occurrence is minimal, they still influence the performance of our models.

4.2. The following tables hold the data on accuracy, macro average and weighted average. Our *performance.txt* file has a comprehensive list of precision, recall and f-1 scores for all classes (emotions and sentiments). As expected, due to the imbalance of the dataset, the accuracy, macro average and weighted average on emotions is very poor.

As for sentiments, the results are better since the distribution is more balanced. This has a direct result on the scores in Table 2 below, with overall higher scores for accuracy, macro average and weighted average. However, the dataset is skewed towards the 'positive' and 'neutral' sentiments, which is reflected in the f-1 scores in the *performance.txt* file.

Due to the very imbalanced dataset, using embeddings as features for part 3, there was no great improvement to the outcome of the models. However, the speed at which it processed the data was much greater.

Table 1: Classifier scores on emotions

Classifier Emotions	Accuracy	Macro Average	Weighted Average
Base-MNB	0.38	0.16	0.31
Base-DT	0.36	0.28	0.36
Base-MLP	0.44	0.24	0.36
Top-MNB	0.39	0.22	0.35
Top-DT	0.38	0.10	0.24
Top-MLP			
Embeddings as Features Base-MLP	0.38	0.11	0.26
Embeddings as Features Top-MLP	0.38	0.11	0.26

Table 2: Classifier scores on sentiments

Classifier Sentiments	Accuracy	Macro Average	Weighted Average
Base-MNB	0.55	0.50	0.54
Base-DT	0.54	0.52	0.55
Base-MLP			
Top-MNB	0.55	0.50	0.54
Top-DT	0.40	0.23	0.31
Top-MLP			
Embeddings as Features Base-MLP	0.51	0.45	0.50
Embeddings as Features Top-MLP	0.52	0.44	0.50

Table 3 and 4 below hold the data from the classification report for the Base Multinomial Naive Bayes classifier for the emotions and sentiments, respectively. Due to the great imbalance in the dataset, certain emotions (grief, nervousness, pride and relief) are not even predicted. The model tends to predict 'neutral' since it is overwhelmingly present. As such the precision of 'neutral' is only 0.37, pointing to a large number of false positives.

As expected, the classification report for sentiments is better, due to a more equal distribution throughout the dataset. The 'positive' sentiment has the best score of the four classes, most likely due to the fact that positive posts are easier to the detect thanks to clear keywords such as: great, awesome and amazing, as well as the use of punctuation ('!'). Furthermore, 'ambiguous' and 'neutral' have lower scores as keywords for these two sentiments are often shared.

Table 3: Classification report Base MNB - emotions

Emotion	Precision	Recall	F1-score	Support
Admiration	0.49	0.45	0.47	2134
Amusement	0.51	0.27	0.35	1233
Anger	0.35	0.12	0.18	1062
Annoyance	0.18	0.06	0.09	1688
Approval	0.25	0.10	0.14	2226
Caring	0.24	0.05	0.08	669
Confusion	0.33	0.06	0.10	965
Curiosity	0.37	0.11	0.17	1143
Desire	0.29	0.02	0.03	439
Disappointment	0.25	0.04	0.07	970

Disapproval	0.24	0.07	0.11	1537
Disgust	0.44	0.09	0.15	585
Embarrassment	0.25	0.01	0.01	303
Excitement	0.28	0.03	0.05	597
Fear	0.20	0.02	0.03	317
Gratitude	0.74	0.68	0.71	1409
Grief	0.00	0.00	0.00	58
Joy	0.32	0.11	0.16	814
Love	0.63	0.35	0.45	1004
Nervousness	0.00	0.00	0.00	187
Neutral	0.37	0.84	0.51	11067
Optimism	0.46	0.14	0.22	914
Pride	0.00	0.00	0.00	157
Realization	0.20	0.03	0.05	972
Relief	0.00	0.00	0.00	156
Remorse	0.51	0.07	0.12	272
Sadness	0.41	0.09	0.15	795
Surprise	0.33	0.07	0.11	691

Table 4: Classification report Base MNB – sentiments

Emotion	Precision	Recall	F1-score	Support
Ambiguous	0.35	0.47	0.40	3771
Negative	0.55	0.60	0.55	7774
Neutral	0.49	0.49	0.49	11067
Positive	0.62	0.70	0.66	11752

Only the classification report for the MNB classifier is analyzed in depth, as analyzing all scenarios would be tedious. All classification reports for the remaining classifiers can be found in the *performance.txt* file.

Another source of low scores is the uneven distribution of classes in our testing and training sets. When using the function call ‘train_test_split’ the dataset is most likely not properly split and results in imbalanced training of the model. For example, the training set can be heavily comprised of the ‘neutral’ emotion, resulting in skewed and improper training for the other emotions in the dataset. When a dataset is very imbalanced, it is recommended to use stratified sampling to ensure relative class frequencies is somewhat preserved in each training and testing sets¹.

¹ https://scikit-learn.org/stable/modules/cross_validation.html#stratification

4.3. Every team member worked on all parts. We would always get on Discord with one person sharing their screen and all communicate our ideas to solve the problems. We would then each contribute to the GitHub repository².

² https://github.com/paulwestenberg/COMP472_MiniProj1