SemEval-2025

BRIDGING THE GAP IN TEXT-BASED EMOTION DETECTION

INTRO TO DEEP LEARNING

Hammad Sajid
Owais Waheed
Muhammad Areeb Kazmi
Kushal Chandani

Final-Evaluation presentation

Problem Statement

Our research addresses the challenge of recognizing perceived emotions in text. The aim is to determine what emotion people believe the speaker may be feeling based on a sentence or short text snippet. By focusing on subtle and complex ways emotions are expressed through language, we highlight the variability in how emotions are perceived and articulated.

Tasks

1

Multi-label Emotion Detection

For each text snippet, the task is to predict the speaker's perceived emotions by assigning labels (0 or 1) to five emotions: joy, sadness, fear, anger, and surprise.

Dataset Format (Task 1)

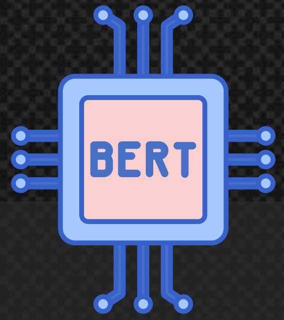
	A	B	C	D	E	F	G
1	id	text	Anger	Fear	Joy	Sadness	Surprise
2	eng_train_track_a_00001	But not very happy.	0	0	1	1	0
3	eng_train_track_a_00002	Well she's not gon na last the whole song like that, so since I'm behind her and the audience can't see below my torso pretty	0	0	1	0	0
4	eng_train_track_a_00003	She sat at her Papa's recliner sofa only to move next to me and start clinging to my arms.	0	0	0	0	0
5	eng_train_track_a_00004	Yes, the Oklahoma city bombing.	1	1	0	1	1
6	eng_train_track_a_00005	They were dancing to Bolero.	0	0	1	0	0
7							
8	eng_train_track_a_00007	But I am exhausted-my eyes feel like they are about to pop out of my head-I need some soothing music and images to help m	0	1	0	1	0
9	eng_train_track_a_00008	We ordered some food at Mcdonalds instead of buying food at the theatre because of the ridiculous prices the theatre has.	1	0	0	0	0
10	eng_train_track_a_00009	Now my parents live in the foothills, and the college is in a large valley.	0	0	0	0	0
11	eng_train_track_a_00010	We get to the porch and my dog starts *growling*, like a big boy growl, like shits going down growl.	0	1	0	0	1
12	eng_train_track_a_00011	I moved my arms, stretching the muscles, watching ribbons of flesh dance around skin and bone.	0	0	1	0	0
13	eng_train_track_a_00012	The room was small but brightly lit and I sat on a two-seater couch facing the counselor across the room, as opposed to loun	0	0	0	0	0
14	eng_train_track_a_00013	The top of the mattress comes up a little above my waist!	0	0	1	0	0
15	eng_train_track_a_00014	I have plenty more.	0	0	1	0	0
16	eng_train_track_a_00015	it took a little longer for my feet to hurt which was nice.	0	0	1	0	0
17	eng_train_track_a_00016	About 2 weeks ago I thought I pulled a muscle in my calf.	0	1	0	1	0
18	eng_train_track_a_00017	I still cannot explain this.	0	1	0	0	1
19	eng_train_track_a_00018	more funny than creepy being on this side of the story:	0	1	1	0	1
20	eng_train_track_a_00019	5 year old me was scarred for life.	0	1	0	1	0
21	eng_train_track_a_00020	The waitress had physical therapy experience and prepared a nice bag of ice for my ankle and the beer was FANTASTIC!	0	0	1	0	0
22	eng_train_track_a_00021	Then I decided to try and get up to go to the restroom, but I couldn't move!	0	1	0	0	1
23	eng_train_track_a_00022	" The cop tells him to have a nice day and walks away.	1	0	1	0	1
24	eng_train_track_a_00023	The following two days, I was in a moderate amount of pain and had very limited range of motion in my arms.	0	1	0	1	0
25	eng_train_track_a_00024	He saw blood and said, "Mommy!	0	1	0	1	1
26	eng_train_track_a_00025	Not the most unnerving feeling, but the most prominent event in my mind is as follows:	0	1	0	0	1
27	eng_train_track_a_00026	When the dust settled I looked over at my wife and saw she was alive and I knew I was alive so I immediately had happiness i	0	0	1	0	0
28	eng_train_track_a_00027	I love you boy.	0	0	1	0	0
29	eng_train_track_a_00028	i brush my teeth at least twice a day.	0	0	0	0	0
30	eng_train_track_a_00029	Needless to say, I turned her down.	0	0	0	0	0

- A train dataset of exact 2769 entries.
- A test dataset of exact 217 entries.

Task 1 - Baseline

Text Loading and Tokenization

Optimized the text for loading and tokenization in order to pass the model.



Fine-tuning

We used Bert-base-uncased (110 M parameters) for the baseline of our task.

Testing the model on dev-dataset

Getting the predictions on the dev-dataset after it has been trained and evaluated.

Task 1 - Baseline

- Hyperparameters

```
optimizer = AdamW(model.parameters(), lr=5e-5, eps=1e-8)

epochs = 4
scheduler = get_linear_schedule_with_warmup(optimizer,
                                            num_warmup_steps=0,
                                            num_training_steps=len(df) * epochs)

training_arguments = TrainingArguments(
    output_dir="./outputs",
    evaluation_strategy="epoch",
    learning_rate=5e-5,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    num_train_epochs=epochs,
    weight_decay=0.01,
    load_best_model_at_end=True,
    metric_for_best_model="f1",
    save_strategy="epoch"
)
```

- Test Scores

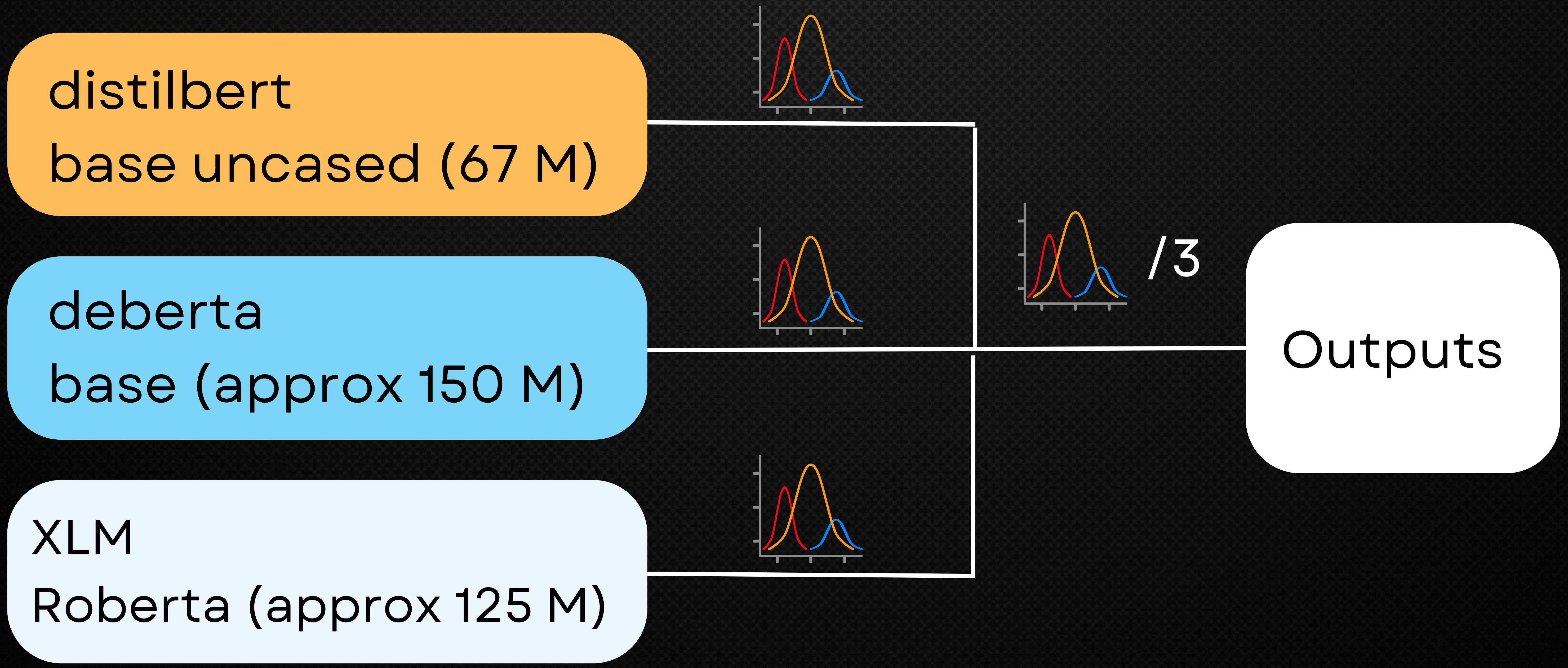
86840	pred_eng_a.zip	2024-10-07 09:17	Finished	0.61

- Train Scores

[556/556 04:13, Epoch 4/4]

Epoch	Training Loss	Validation Loss	Accuracy	F1	Auc
1	No log	0.393659	0.366426	0.712211	0.795974
2	No log	0.370237	0.404332	0.726457	0.803856
3	No log	0.367175	0.449458	0.737580	0.808690
4	0.294300	0.376198	0.460289	0.745434	0.816454

Task 1 - Ensemble Approach



Task 1 - Ensemble Results

Model	Epoch	Training Loss	Validation Loss	Accuracy	F1	AUC
DistilBERT (0.82)	7	0.034400	0.365389	0.614692	0.835473	0.874029
DeBERTa (0.76)	8	1.016900	1.714891	0.458484	0.798921	0.867681
XLM-Roberta (0.81)	7	0.055300	0.370612	0.593261	0.826517	0.873846

- Test Scores (increased from 0.61 -> 0.68)

92327	pred_eng_b.zip	2024-10-17 14:03	Finished	0.68

Task 1 - Data Augmentation

Gemini

- Base Augmentation generating to around 2000 rows of data.
- Limits: API exhausting and runtime disconnects

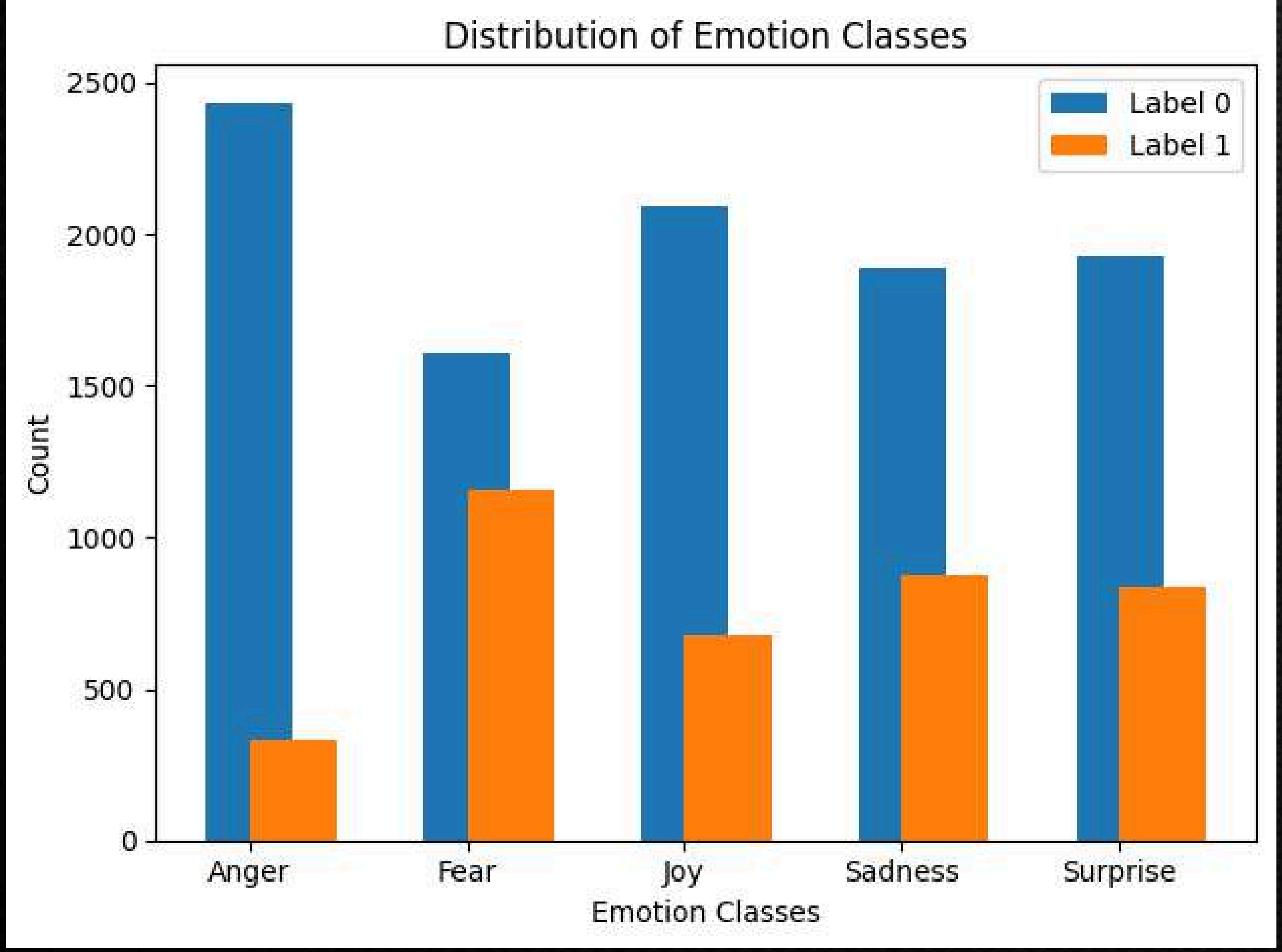
OpenAI

ChatGPT 3.5

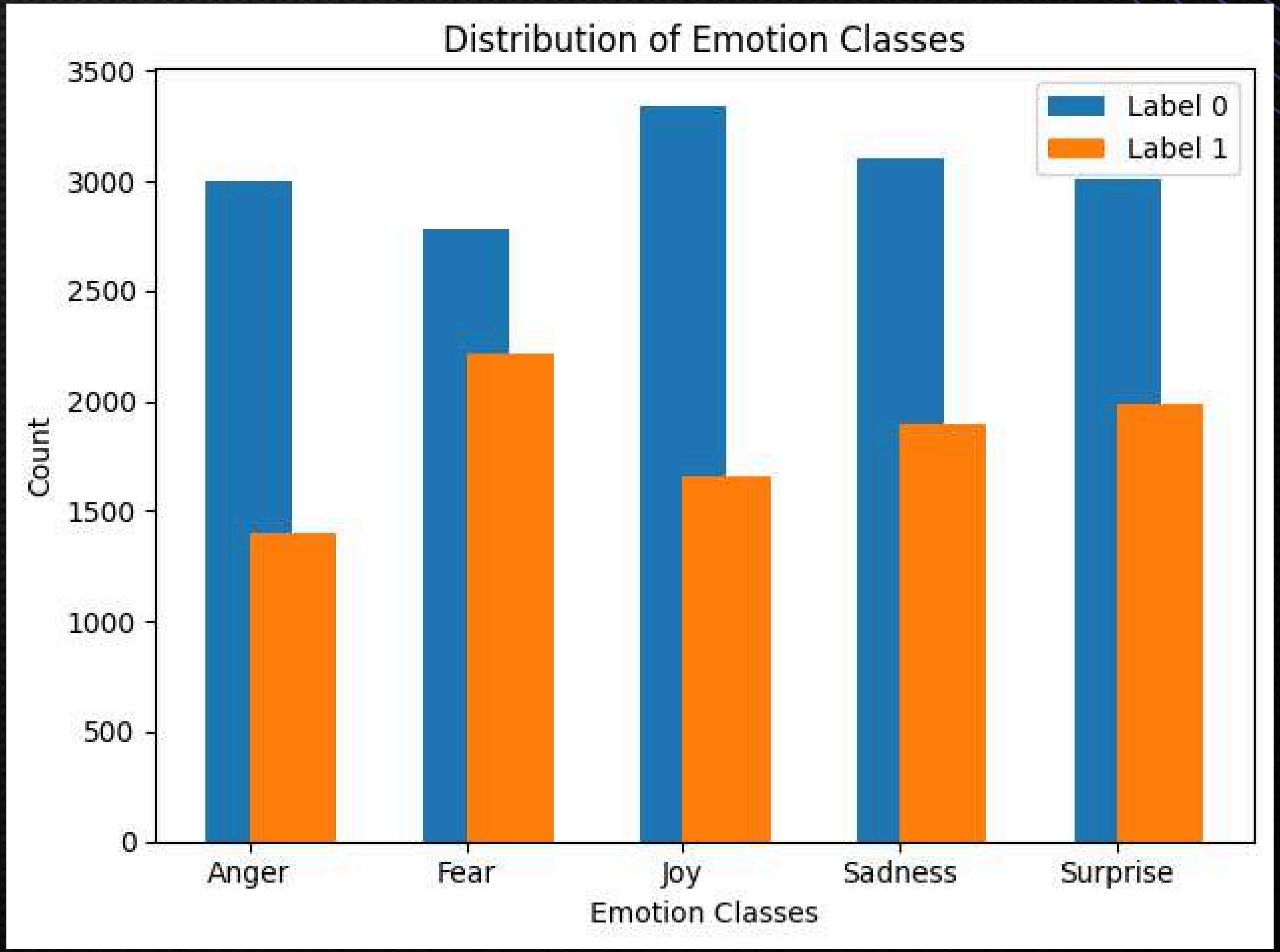
- Main Augmentation generating to around 8400 rows of data.
- Custom prompting and quick results. No limitations.

Task 1 - Data Augmentation

Unaugmented



Augmented



Task 1 – Ensemble + Data Augmentation

Model	Epoch	Training Loss	Validation Loss	Accuracy	F1	AUC
XLM-RoBERTa	8	0.037500	0.257887	0.734000	0.902821	0.921191
DeBERTa	8	1.753800	2.023195	0.623000	0.866981	0.900965
DistilBERT	8	0.020900	0.243800	0.719000	0.897938	0.915951

- Test Scores (increased from 0.61 -> 0.68 -> 0.73)
  currently among top 4 scores out of 63 submissions

166451	pred_eng_a.zip	2024-11-21 13:01	Finished	0.73

Tasks

1

Multi-label Emotion Detection

For each text snippet, the task is to predict the speaker's perceived emotions by assigning labels (0 or 1) to six emotions: joy, sadness, fear, anger, surprise, and disgust.

2

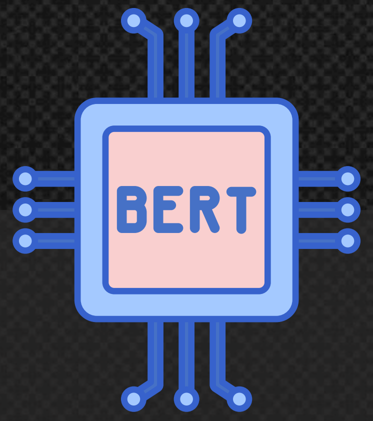
Multi-label Cross-lingual Emotion detection

For each text snippet **in one language**, the task is to predict the speaker's perceived emotions **in another language** by assigning labels (0 or 1) to six emotions: joy, sadness, fear, anger, surprise, and disgust.

Task 2 - Cross-lingual Emotion Detection - Baseline

Tokenization

Text preprocessing and tokenization and pass it to Bert tokenizer.



Fine Tuning

Split the train data into 80/20 train and test sets and pass it to the bert-multilingual-cased model for fine-tuning.

Testing on Different language

Now we will test the fine-tuned model on different languages except English as it does not have one column (Disgust)

Dataset Format (Task 2)

	id	text	Anger	Disgust	Fear	Joy	Sadness	Surprise
0	deu_train_track_a_00001	Nein nein nix da, fuck den schön weiter ab bis...	1	0	0	0	0	0
1	deu_train_track_a_00002	Vor 100 Jahren ging man auf die Strasse weil m...	1	1	0	0	0	0
2	deu_train_track_a_00003	Wann endlich Gabby Epstein heiraten	0	0	0	0	0	0
3	deu_train_track_a_00004	Wegen dem Song 1001 Nacht, dem einen Phantomto...	0	0	0	1	0	0
4	deu_train_track_a_00005	Dann sollten wir versuchen, sie zu erreichen. ...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
2598	deu_train_track_a_02599	Das nenne ich guten Geschmack, mein Geologenhe...	0	0	0	1	0	0
2599	deu_train_track_a_02600	Zeigt mir irgendwie, dass es auch bei den Nörg...	0	0	0	1	0	0
2600	deu_train_track_a_02601	1.) Ist es möglich dein Tätigkeitsbereich etwa...	0	0	0	0	0	0
2601	deu_train_track_a_02602	Egal nicht alles muss man unter Kontrolle haben	0	0	0	0	0	0
2602	deu_train_track_a_02603	Klar, das ist der Grund. Auf keinen Fall die P...	1	1	0	0	0	0

2603 rows × 8 columns

- A train dataset of exact 2603 entries.
- A test dataset of exact 401 entries.

Task 2 - Baseline Scores

Model	Epoch	Training Loss	Validation Loss	Accuracy	F1	AUC
FacebookAI/xlm-roberta-base	4	0.340200	0.297479	0.412371	0.270096	0.580038
distilbert-base-multilingual	4	0.315600	0.278888	0.478645	0.436782	0.669427
mBERT	4	0.340000	0.300690	0.382916	0.192771	0.551868

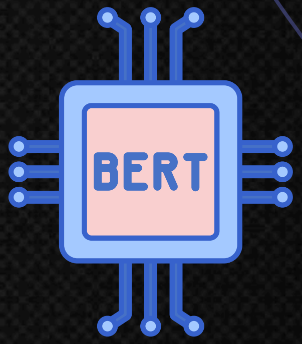
Task 2 - Cross-lingual Emotion Detection - Using Translation

Russian

Target Language

Translated to English

By University of Helsinki's Russian to English Translation model



Model FineTuning

German

Training Language

Translated to English

By University of Helsinki's German to English Translation model

Emotion Detection Testing

Task 2 - Post-Translation Scores

Model	Epoch	Training Loss	Validation Loss	Accuracy	F1	AUC
FacebookAI/xlm-roberta-base	4	No log	0.380091	0.404990	0.551111	0.724607
distilbert-base-multilingual	4	No log	0.357402	0.420345	0.579285	0.733115
mBERT	4	0.323200	0.371087	0.351248	0.577624	0.733115

Tasks

1

Multi-label Emotion Detection

For each text snippet, the task is to predict the speaker's perceived emotions by assigning labels (0 or 1) to six emotions: joy, sadness, fear, anger, surprise, and disgust.

2

Multi-label Cross-lingual Emotion detection

For each text snippet **in one language**, the task is to predict the speaker's perceived emotions **in another language** by assigning labels (0 or 1) to six emotions: joy, sadness, fear, anger, surprise, and disgust.

3

Emotion Intensity

Given a target text and a target perceived emotion, predict the intensity for each class as 0 for no emotion, 1 for low, 2 for moderate, and 3 for high degree of emotion.

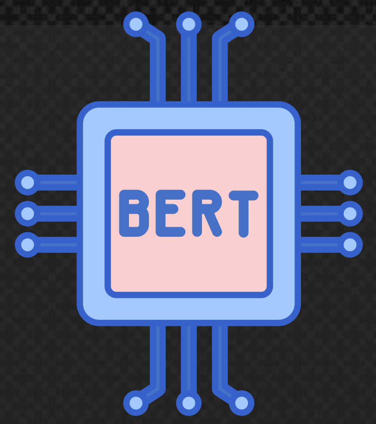
Dataset Format (Task 3)

id	text	Anger	Fear	Joy	Sadness	Surprise
eng_dev_track_b_00001	I noticed this months after moving in and d	0	1	0	0	0
eng_dev_track_b_00002	can't wait to be in another wedding!	0	0	3	0	0
eng_dev_track_b_00003	Just getting out of the house put a smile or	0	0	3	0	0
eng_dev_track_b_00004	Your sister is a heavy sleeper.	0	1	0	0	0
eng_dev_track_b_00005	The police were called.	0	2	0	0	1
eng_dev_track_b_00006	We moved to a Labor and Delivery Room a	0	1	0	1	0
eng_dev_track_b_00007	I was beyond furious.	2	1	0	0	0
eng_dev_track_b_00008	(you see, i have these afi patches that i bo	0	0	0	0	0
eng_dev_track_b_00009	My heart sank.	0	0	0	3	0
eng_dev_track_b_00010	It was late and they were asleep.	0	0	0	0	0
eng_dev_track_b_00011	What is it about this winter that is making r	0	2	0	1	0
eng_dev_track_b_00012	The doors were getting ready to close, and	0	0	0	0	0
eng_dev_track_b_00013	<< < using my brain?	0	1	0	0	1
eng_dev_track_b_00014	and my feet hurt.	0	1	0	0	0
eng_dev_track_b_00015	After standing in a few times in the past his	0	1	0	1	1
eng_dev_track_b_00016	The way I was positioned with the air movi	0	2	0	0	0
eng_dev_track_b_00017	... but Raul never called.	1	1	0	1	1
eng_dev_track_b_00018	"I don't have time to arrest drunk drivers b	1	3	0	1	0
eng_dev_track_b_00019	My heart was beating fast from excitement	0	1	2	0	0
eng_dev_track_b_00020	I untie my bear bag and grab some food ins	0	0	1	0	0
eng_dev_track_b_00021	Longest, most awkward drive I've ever take	0	2	0	1	0
eng_dev_track_b_00022	It has sat with me, back in the pit of my sto	0	1	0	1	0
eng_dev_track_b_00023	I know not why, I wipe my face.	0	1	0	1	0
eng_dev_track_b_00024	It's so stressful but I'd never wanna be with	0	1	0	1	0
eng_dev_track_b_00025	My sister was walking backwards and bump	0	1	0	0	0

Task 3 - Baseline

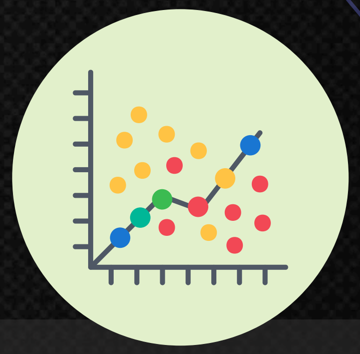
Data Tokenization

Used Bert tokenizer for the tokenization of the data



Bert Pretrained Model

Used a pre-trained BERT model to generate logits, with Mean Squared Error (MSE) as the loss function instead of cross-entropy.



Linear Regressor

Applied a linear regression layer to transform the logits into the desired intensity range of 0-3 by rounding off the outputs.

Task 3 - Baseline Scores

Your results for eng track B are:
Anger: 0.6769
Fear: 0.5704
Joy: 0.7132
Sadness: 0.7821
Surprise: 0.6717
Average Pearson r: 0.6829

Task 3 - Ensemble Approach

distilbert
base uncased (67 M)

deberta-v3-large
(304M)

deberta
base (approx 150 M)

bert-base-uncased
(110 M)

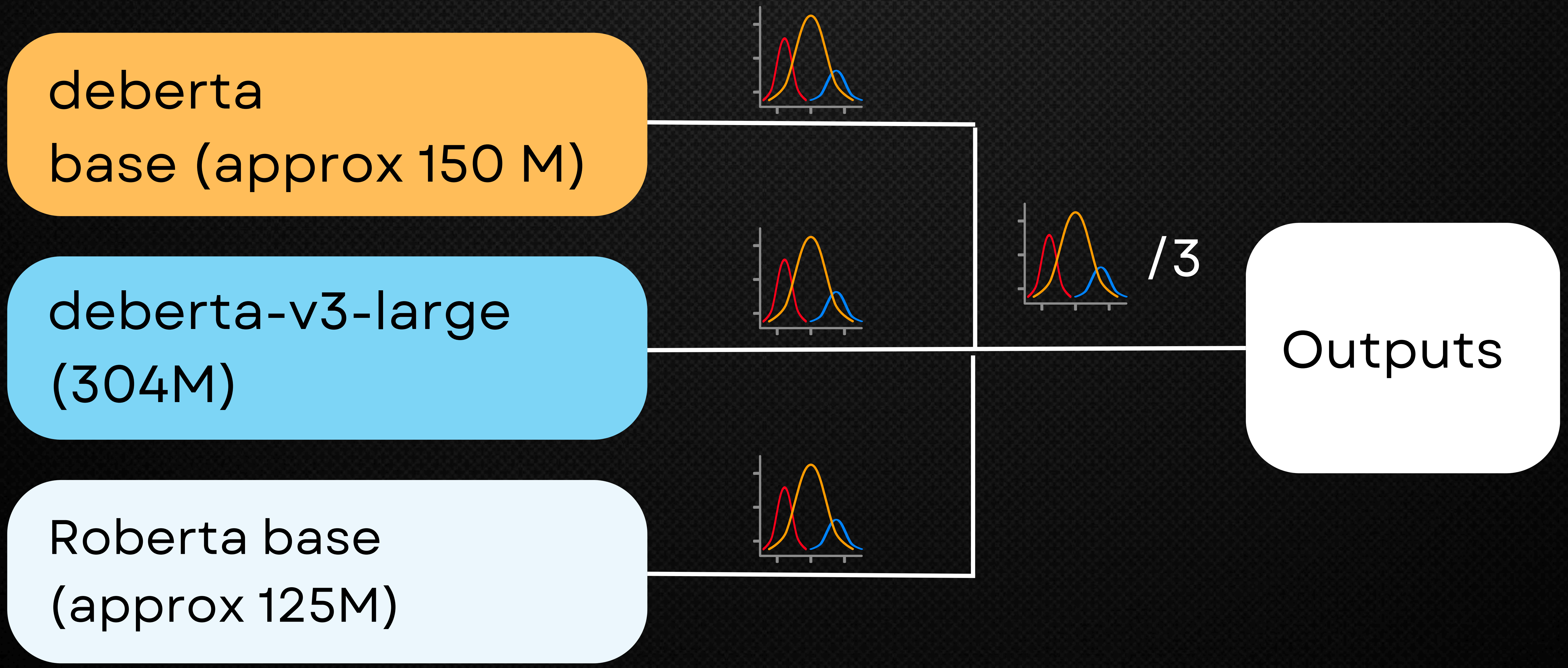
XLM Roberta
(approx 279M)

Roberta base
(approx 125M)

Task 3 - Ensemble Approach

Model	Epoch	Training Loss	Validation Loss	MSE
microsoft/deberta-v3-base	5	0.120200	0.225958	**0.225958**
microsoft/deberta-v3-large	5	0.046600	0.233142	**0.233142**
roberta-base	5	0.086100	0.257203	**0.257203**
bert-base-uncased	4	0.074200	0.271673	**0.271673**
distilbert-base-uncased	3	0.135600	0.292721	**0.292721**
FacebookAI/xlm-roberta-base	3	0.404500	0.363043	**0.363043**

Task 3 - Ensemble Approach



Task 3 - Ensemble Scores

Your results for eng track B are:
Anger: 0.7201
Fear: 0.6403
Joy: 0.7154
Sadness: 0.835
Surprise: 0.7287
Average Pearson r: 0.7279

Test Scores (increased from 0.68 -> 0.73)

Task 3 - Data Augmentation limitations

OpenAI

ChatGPT 3.5

- Main Augmentation generated around 8000 rows of data.
- Active to passive and synonyms techniques was used however the scores were reduced.
- Intensity wasn’t captured entirely.

References

[1] A. Rajesh, S. A. Abirami, A. C. Chandrabose, and S. Kumar, "SSN_Semeval10 at SemEval-2024 Task 10: Emotion Discovery and Reasoning its Flip in Conversations," Proc. 18th Int. Workshop SemEval-2024, pp. 553-557, June 2024.

[2] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," Computing Research Repository, 2013. [Online]. Available: http://arxiv.org/abs/1308.6297

[3] S. Poria, N. Majumder, R. Mihalcea, and E. Hovy, "Emotion recognition in conversation: Research challenges, datasets, and recent advances," IEEE Trans. Knowl. Data Eng., vol. 28, no. 2, pp. 496-509, 2016.

[4] A. Wadhawan and A. Aggarwal, "Towards emotion recognition in Hindi-English code-mixed data: A transformer based approach," in Proc. Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Online, 2021, pp. 195-202.

[5] Y. Wang, Y. Li, P. Liang, L.-P. Morency, P. Bell, and C. Lai, "Cross-attention is not enough: Incongruity-aware dynamic hierarchical fusion for multimodal affect recognition," IEEE Access, vol. 11, pp. 13583-13593, 2023.

[6] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, "Sentiment embeddings with applications to sentiment analysis," IEEE Trans. Knowl. Data Eng., vol. 28, no. 2, pp. 496-509, 2016.

[7] R. Pan, J. A. García-díaz, D. Roldán, and R. Valencia-garcía, "UMUTeam at SemEval-2024 Task 10: Discovering and Reasoning about Emotions in Conversation using Transformers," Proc. 18th Int. Workshop SemEval-2024, pp. 703-709, June 2024.

Thank You