Text Based Emotion Detection

SemEval2025-Task 11 [1]

CS779
Group-3

Tafazzul Nadeem(232110401)
Riyansha Singh (232110601)
Suyamoon Pathak(231110091)

Supervisor: Dr. Ashutosh Modi

Indian Institute of Technology, Kanpur

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED
- Previous works on TBED

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED
- Previous works on TBED
- Shortlisted Models

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED
- Previous works on TBED
- Shortlisted Models
- Experiments with shortlisted models

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED
- Previous works on TBED
- Shortlisted Models
- Experiments with shortlisted models
- Results of the provided datasets on those models

Agenda

- Intro to the SemEval Task along with it's tracks (A,B,C)
- Datasets provided by the coordinator
- Available Datasets on TBED
- Previous works on TBED
- Shortlisted Models
- Experiments with shortlisted models
- Results of the provided datasets on those models
- Future Directions

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker
- Five emotions for English - joy, sadness, fear, anger, surprise

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker
- Five emotions for English - joy, sadness, fear, anger, surprise
- Three Tracks:

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker
- Five emotions for English - joy, sadness, fear, anger, surprise
- Three Tracks:
  - Track A: Multi-label Emotion Detection

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker
- Five emotions for English - joy, sadness, fear, anger, surprise
- Three Tracks:
  - Track A: Multi-label Emotion Detection
  - Track B: Emotion Intensity

SemEval 2025 - Task 11

- Emotion Detection: Determining what emotion most people will think the speaker may be feeling given a sentence or a short text snippet uttered by the speaker
- Five emotions for English - joy, sadness, fear, anger, surprise
- Three Tracks:
  - Track A: Multi-label Emotion Detection
  - Track B: Emotion Intensity
  - Track C: Cross-lingual Emotion Detection

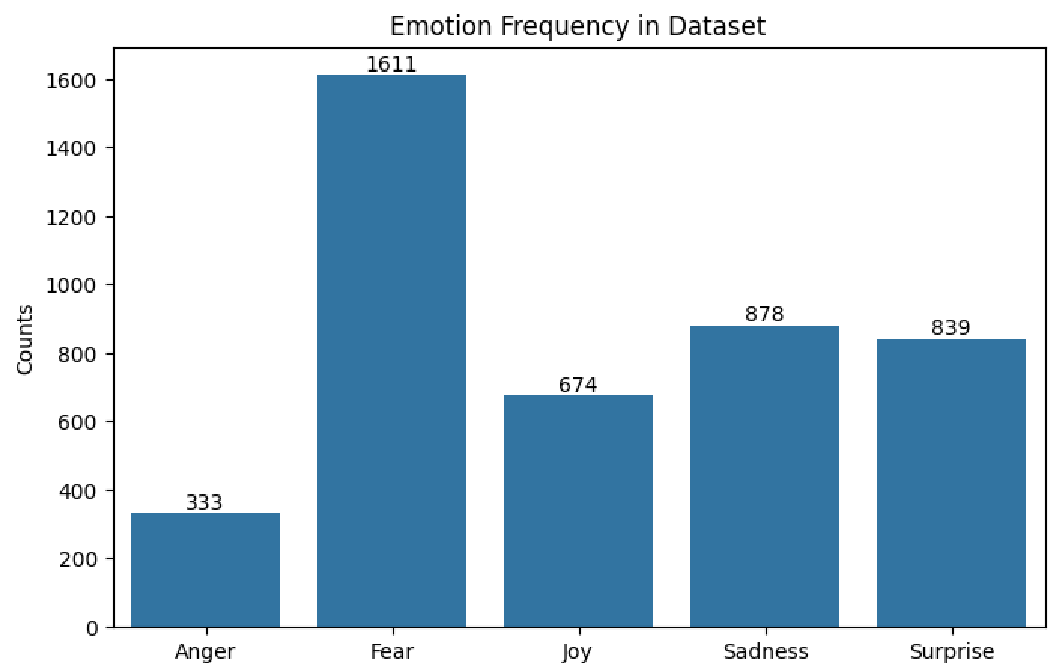
Dataset for Track A

	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 l	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	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	eng_train_track_a_00006	Still had sex with her, though.	0	0	1	0	0
8	eng_train_track_a_00007	But I am exhausted-my eyes feel like they are about to pop out o	0	1	0	1	0
9	eng_train_track_a_00008	We ordered some food at Mcdonalds instead of buying food at th	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	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	0	1	0	0	1
12	eng_train_track_a_00011	I moved my arms, stretching the muscles, watching ribbons of fl	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 cou	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 ni	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 coul	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	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	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	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

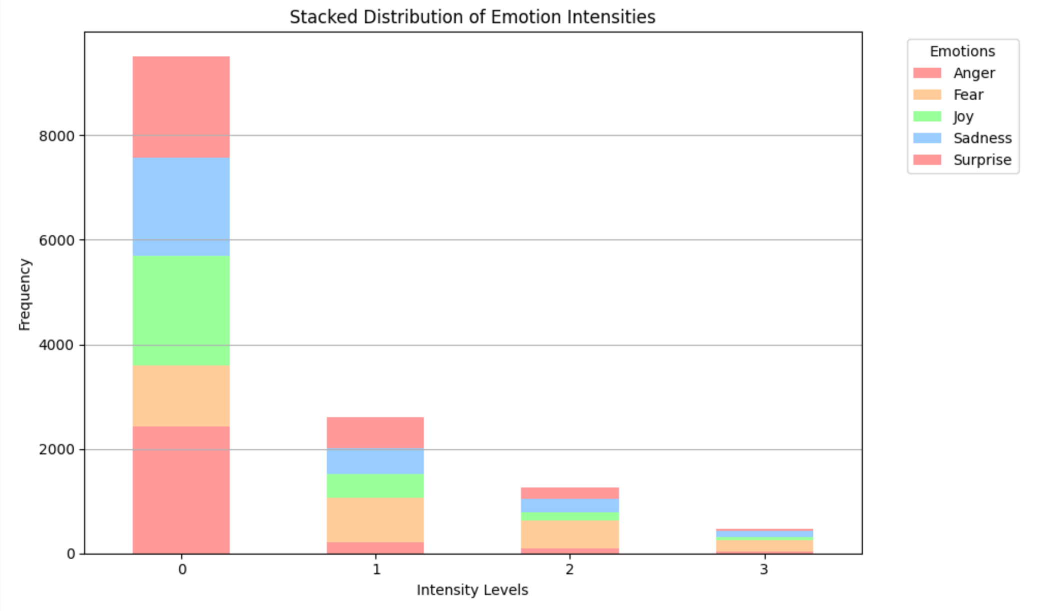
Dataset for Track B

	B	C	D	E	F	G
1	text	Anger	Fear	Joy	Sadness	Surprise
2	My Spanish language skills were fairly basic.	0	1	0	0	0
3	Don't mess with my orange juice.	2	0	0	0	0
4	So, I am from a science background and analyze everything skeptically.	0	0	0	0	0
5	I was writing away.	0	0	1	0	0
6	Apparently it wasn't as life threatening as I believed at the time because the doctor c	0	1	0	1	0
7	I could not believe how fast it went considering I had to stay on my back the whole la	0	1	0	0	1
8	The room really isn't a room, it is as though she is kneeling on the floor but I don't se	0	2	0	0	2
9	It looks like dark smelly applesauce.	0	1	0	0	1
10	like seriously, alot can not understand and im slow at registering words in my brain c	0	2	0	1	0
11	Life readings obviously defective.	1	1	0	0	0
12	months before that, i just prayed to God that i wanted to meet someone right for me,	0	0	0	2	0
13	My grandmother fell very ill this past summer, and ended up suffering from encepha	0	1	0	3	0
14	A highly intoxicated roommate and a black girl going at it.	1	2	0	0	2
15	You are the demon.	2	1	0	0	1
16	I had an ant crawling on my elbow just now, and I flicked him off.	0	1	0	0	0
17	Immediately my throat tightens.	0	3	0	0	0
18	Kick my heels up and shout.	0	0	2	0	0
19	I closed my eyes as I took in a deep breath.	0	0	1	0	0
20	he got pretty fucking scared and started sprinting away from me.	0	3	0	0	1
21	Nothing was touching it.	0	1	0	0	1
22	Met a lot of random canadians that night.	0	0	1	0	1
23	I saw that, too.	0	0	0	0	1
24	just the thought of having to get up and speak in front of 30 people makes my heart	0	3	0	0	0
25	Moral of the story?	0	0	0	0	1
26	I got my first raspberry from a crowd surfer falling and my face hitting the monitor, ba	0	1	0	1	1
27	At this point he decided to consult the woman at the front desk for more information.	0	1	0	0	0
28	When she decided it was time to go back to my stall, I didn't want to leave and I pre	0	1	0	1	0
29	Finding them, I whipped one out and put it in my ear, swabbing.	0	0	0	0	0
30	I looked down only to see a flashing red light buried in the dirt under my feet.	0	1	0	0	1

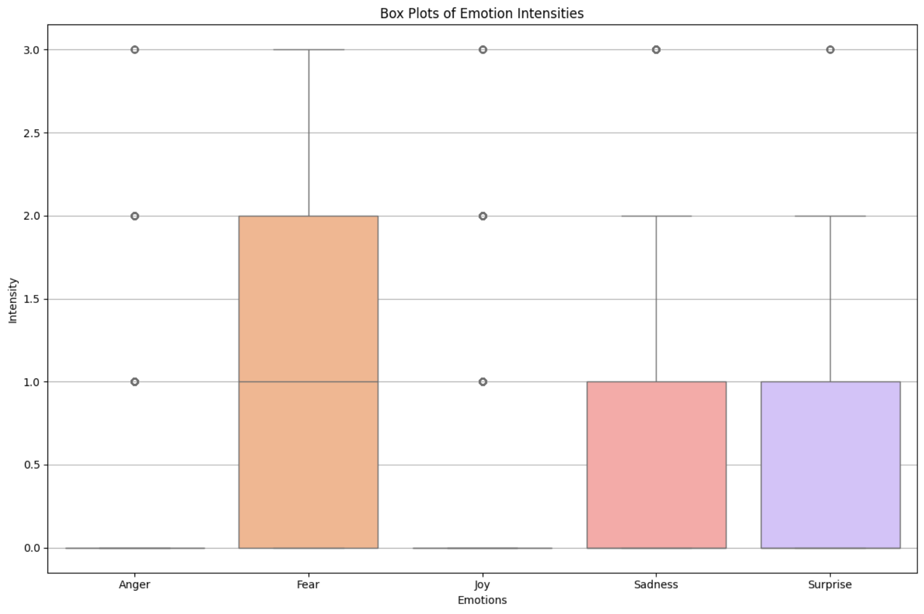
How are the emotions distributed?



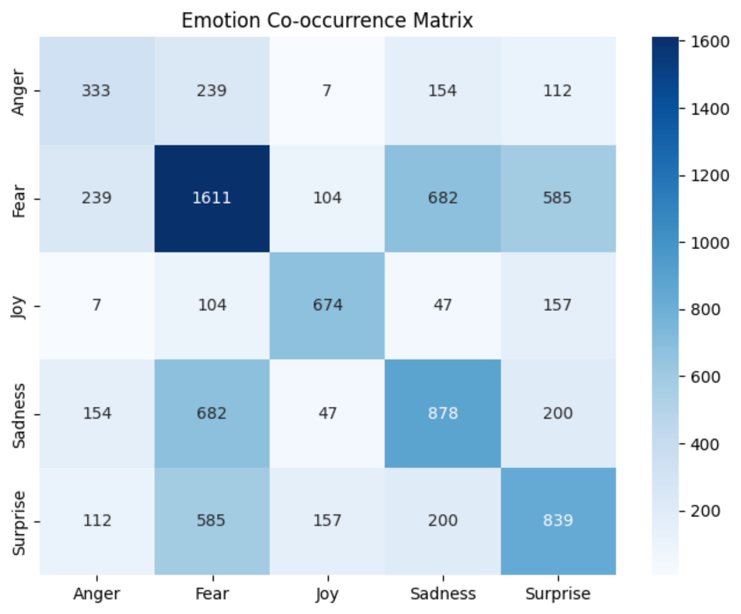
How are the intensity levels distributed?



How are the intensity levels distributed?



How frequently are the emotions co-occurring?



Datasets available for TBED

Dataset	Source	Size	Emotion Model
GoEmotions [8]	Blogs (Reddit) ; labeled between 27 emotional classes.	58 K	Discrete, Componential
ISEAR (2017) [19]	Worldwide survey on causes and consequences. Emotional state reported as happiness, fear, anger, sorrow, disgust, humiliation, and guilt.	7.6K	Dimensional
SemEval (2018) [15]	Blogs, Tweets	23 K	Discrete
SemEval (2019) [6]	Text conversation between two people. Dialogue classified as either happy, angry, unhappy, or other	39 K	Discrete
EMOBANK [5]	Blogs, newspaper headlines, letters, and travel guides	10 K	Dimensional
SuperTweetEval [3]	Benchmark for evaluating tweet-level sentiment and language. It contains a diverse collection of tweets, annotated for 12 heterogeneous NLP tasks.		Discrete; Dimensional
Valence and arousal [17]	Facebook posts. Contains labels for two different grades, which are not related to each other:	3.2K	Dimensional
CBET [10]	Tweets	77k	Componential
ISEAR (2018)	Cross-cultural studies	40 K	Discrete

SpanEmo: Casting Multi-label Emotion Classification as Span-prediction

- Proposes a new model called "SpanEmo" [2] that predicts the emotion classes directly from the label set, allowing it to capture relationships between words and the corresponding emotions.

SpanEmo: Casting Multi-label Emotion Classification as Span-prediction

- Proposes a new model called "SpanEmo" [2] that predicts the emotion classes directly from the label set, allowing it to capture relationships between words and the corresponding emotions.
- Introduces a novel loss function Label-correlation aware loss (LCA) for handling co-existing emotions.

$$\mathcal{L}_{LCA}(y, \hat{y}) = \frac{1}{|y_0||y_1|} \sum \exp(\hat{y}_p - \hat{y}_q) \tag{1}$$

SpanEmo: Casting Multi-label Emotion Classification as Span-prediction

- Proposes a new model called "SpanEmo" [2] that predicts the emotion classes directly from the label set, allowing it to capture relationships between words and the corresponding emotions.
- Introduces a novel loss function Label-correlation aware loss (LCA) for handling co-existing emotions.

$$\mathcal{L}_{LCA}(y, \hat{y}) = \frac{1}{|y_0||y_1|} \sum \exp(\hat{y}_p - \hat{y}_q) \tag{1}$$

- Experiments on **SemEval2018 multi-label emotion data** [15].

SpanEmo: Casting Multi-label Emotion Classification as Span-prediction

- Proposes a new model called "SpanEmo" [2] that predicts the emotion classes directly from the label set, allowing it to capture relationships between words and the corresponding emotions.
- Introduces a novel loss function Label-correlation aware loss (LCA) for handling co-existing emotions.

$$\mathcal{L}_{LCA}(y, \hat{y}) = \frac{1}{|y_0||y_1|} \sum \exp(\hat{y}_p - \hat{y}_q) \tag{1}$$

- Experiments on **SemEval2018 multi-label emotion data** [15].
- Performed over three languages: English, Arabic, Spanish.

Cross-lingual Emotion Detection through Large Language Models

- Leveraged a mix of open-source and proprietary Large Language Models (LLMs), including GPT-4, Claude-Opus, LLAMA-3-8B, Gemma-7B, and Mistral-v2-7B.

Cross-lingual Emotion Detection through Large Language Models

- Leveraged a mix of open-source and proprietary Large Language Models (LLMs), including GPT-4, Claude-Opus, LLAMA-3-8B, Gemma-7B, and Mistral-v2-7B.
- Models were used in zero-shot configurations or fine-tuned, with precision variations (4-bit and 16-bit) tested for computational efficiency and performance.

Cross-lingual Emotion Detection through Large Language Models

- Leveraged a mix of open-source and proprietary Large Language Models (LLMs), including GPT-4, Claude-Opus, LLAMA-3-8B, Gemma-7B, and Mistral-v2-7B.
- Models were used in zero-shot configurations or fine-tuned, with precision variations (4-bit and 16-bit) tested for computational efficiency and performance.
- Stochastic weight averaging (SWA) in this case only led to a minor improvement and tech-niques like augmentation using other datasets did not improve performance

Shortlisted Models

- Small Models:
  - SamLowe/roberta-base-go_emotions [14]
  - cardiffnlp/twitter-roberta-large-emotion-latest [3]
  - Emanuel/twitter-emotion-deberta-v3-base [12]
- Large Models:
  - meta-llama/Meta-Llama-3-8B-Instruct [9]

SamLowe/roberta-base-go-emotions

- Pretrained Model: FacebookAI/roberta-base[13]
- Linear classifier trained on top of Roberta base model
- Dataset trained on: google-research-datasets/go-emotions [8]
- Results[14] on Goemotions dataset:
  - Accuracy: 0.474
  - F1: 0.450
- 28 classes for classification (multi-emotion labels)

cardiffnlp/twitter-roberta-large-emotion-latest

- Pretrained model: FacebookAI/roberta-large[13]
- Linear classifier finetuned on: 154M tweets
- Finetuned on: cardiffnlp/super_tweeteval[3]
- Results[3] on super_tweeteval dataset:
  - Avg Macro F1: 0.58
- Results[18] on further finetuned on Goeomotions dataset:
  - F1 : 0.92
- 11 classes for classification (multi-emotion labels)

Emanuel/twitter-emotion-deberta-v3-base

- Pretrained model: microsoft/deberta-v3-base[11]
- Pretrained on: 160GB Data
- Linear classifier finetuned on: dair-ai/emotion[7]
- Results[12] on dair-ai/emotion dataset:
  - Accuracy: 0.937
- 6 classes for classification (single-emotion labels)

meta-llama/Meta-Llama-3-8B-Instruct

- Pretrained Model: LLama3-8B[9]
- Trained on: 15T+ web tokens
- instruction tuned

Experiments

- The training data was released on 10/09/24
- The models were evaluated on the training data for english language without any fine-tuning
- Accuracy, Micro, Macro and Weighted F-1 Scores were calculated for the predictions of these models for Track A dataset
- Accuracy calculated for Track B dataset on the same models

Track A: Multilabel emotion detection

- Full Fine-tuning of Off-the-Shelf Models
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.43
- Micro F1 score: 0.60
- Macro F1 score: 0.49

Track A: Multilabel emotion detection

- Full Fine-tuning of Off-the-Shelf Models
- Trained on : SamLowe/roberta-base-go_emotions
- Multi-label accuracy (Jaccard score): 0.44
- Micro F1 score: 0.61
- Macro F1 score: 0.49

Track A: Multilabel emotion detection

- Training of an Added Classifier Layer Only
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.57
- Micro F1 score: 0.73
- Macro F1 score: 0.69

Track A: Multilabel emotion detection

- Training of an Added FC Layer Along with the Last Classifier Layer
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.62
- Micro F1 score: 0.77
- Macro F1 score: 0.75

Track A: Multilabel emotion detection

- Training the current best model on Entailment Approach
- The dataset is converted to Premise-Hypothesis pair dataset with a label 0 or 1 for Hypothesis being in contradiction/neutrality of the Premise and 1 for Hypothesis being entailment of the Premise respectively for every emotion:
- **Original Sample:**
  Text: But not very happy.
  Labels: [0.0, 0.0, 1.0, 1.0, 0.0] ('Anger','Fear','Joy','Sadness','Surprise')
- **Converted to:**
  premise: But not very happy.
  Hypothesis: The speaker is feeling Anger
  Labels: [0.0] (Neutral or Contradiction)
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.60
- Micro F1 score: 0.73
- Macro F1 score: 0.73

Track A: Multilabel emotion detection

- Oversampling minority class with replacement- Method-1
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.59
- Micro F1 score: 0.73
- Macro F1 score: 0.73

Track A: Multilabel emotion detection

- Oversampling minority class with replacement- Method-2
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.64
- Micro F1 score: 0.78
- Macro F1 score: 0.77

Track B: Emotion Intensity Detection

- Oversampling minority class with replacement- Method-2
- Trained on : cardiffnlp/twitter-roberta-large-emotion-latest
- Multi-label accuracy (Jaccard score): 0.64
- Micro F1 score: 0.78
- Macro F1 score: 0.77

Track C: Cross lingual emotion detection

- Pretrained Model: cardiffnlp/twitter-xlm-roberta-base-sentiment [4]
- trained on 198M tweets and finetuned for sentiment analysis done on 8 languages
- **Zero-Shot Transfer:** fine-tuned on Russian text and evaluated on English text.
- Multi-label accuracy (Jaccard score):**0.10**,
  Micro F1 score: **0.17**,
  Macro F1 score: **0.13**.

Employing External Dataset

- The XED dataset [16] is evaluated for the purpose of augmentation.
- The dataset consists of emotion annotated movie subtitles from OPUS. Plutchik's 8 core emotions are used to annotate.

Number of annotations	**24164 + 9384 neutral**
Number of unique data points	17530 + 6420 neutral
Number of emotions	8 (+pos, neg, neu)
Number of annotators	108 (63 active)

Table 1: Statistics for XED

Analysis

- Similarity score between the two datasets was computed to be 0.26. While not highly similar, the overlap suggests potential for useful augmentation.
- The XED dataset was utilized for fine tuning the best-performing model **cardiffnlp/twitter-roberta-large-emotion-latest**
- The model trained achieved comparable accuracy to the results obtained using the original dataset on training with a F1-score of **0.540793**.

Results

Model Name	Accuracy	Micro F1	Macro F1	Weighted F1
SamLowe/roberta-base-go_emotions	0.21	0.45	0.44	0.42
cardiffnlp/twitter-roberta-large-emotion-latest	**0.29**	0.54	0.53	0.50
Emanuel/twitter-emotion-deberta-v3-base	0.16	0.45	0.40	0.45
meta-llama/Meta-Llama-3-8B-Instruct	0.24	**0.58**	**0.59**	**0.58**

Table 2: Results for Track A dataset without fine-tuning on the shortlisted transformer models

Model Name	Accuracy
SamLowe/roberta-base-go_emotions	0.11
cardiffnlp/twitter-roberta-large-emotion-latest	0.10
meta-llama/Meta-Llama-3-8B-Instruct	**0.18**

Table 3: Results for Track B dataset without fine-tuning on the shortlisted transformer models

Results

Model Name	Track	Accuracy	Micro F1	Macro F1	Weighted F1
SamLowe/roberta-base-go_emotions	A	0.21	0.45	0.44	0.42
cardiffnlp/twitter-roberta-large-emotion-latest	A	0.29	0.54	0.53	0.50
Emanuel/twitter-emotion-deberta-v3-base	A	0.16	0.45	0.40	0.45
meta-llama/Meta-Llama-3-8B-Instruct	A	0.24	0.58	0.59	0.58
cardiffnlp/twitter-roberta-large-emotion-latest (Full Fine-tuning)	A	0.43	0.60	0.49	-
SamLowe/roberta-base-go_emotions (Full Fine-tuning)	A	0.44	0.61	0.49	-
cardiffnlp/twitter-roberta-large-emotion-latest (Added Classifier Layer Only)	A	0.57	0.73	0.69	-
cardiffnlp/twitter-roberta-large-emotion-latest (Added FC Layer Along with the Last Classifier Layer)	A	0.62	0.77	0.75	-
cardiffnlp/twitter-roberta-large-emotion-latest (Entailment Approach)	A	0.60	0.73	0.73	-
cardiffnlp/twitter-roberta-large-emotion-latest (Oversampling minority class with replacement- Method-1)	A	0.59	0.73	0.73	-
cardiffnlp/twitter-roberta-large-emotion-latest (Oversampling minority class with replacement- Method-2)	A	**0.64**	**0.78**	**0.77**	-
After Data Augmentation					
SamLowe/roberta-base-go_emotions	A	0.40	0.59	0.48	0.55
cardiffnlp/twitter-roberta-large-emotion-latest	A	0.61	0.76	0.74	0.72
Emanuel/twitter-emotion-deberta-v3-base	A	0.24	0.49	0.44	0.45
cardiffnlp/twitter-xlm-roberta-base-sentiment	C	**0.10**	**0.17**	**0.13**	**0.15**

Results for Track A , B and C

Results

Model Name	Anger (Pear R.)	Fear (Pear R.)	Joy (Pear R.)	Sadness (Pear R.)	Surprise (Pear R.)	Average Pearson r
SamLowe/roberta-base-go_emotions	A	0.21	0.45	0.44	0.42	

Results for Track B

Data Augmentation

- Dataset Limitation: The provided dataset by SemEval was relatively small, consisting of only 2,768 training samples, prompting the need for data augmentation to improve model performance.
- Data Augmentation Strategy: ChatGPT 4o was used to generate synthetic data. From the original dataset, 100 samples were randomly selected, and 50 new entries were created based on these samples.

Prompt Used for Data Augmentation

Prompt: `You are an AI model specializing in text-based emotion detection. I need you to generate synthetic data in the following format:`

`Data Structure: Each entry should consist of a line of text, followed by binary labels for five emotions: Anger, Fear, Joy, Sadness, and Surprise.`

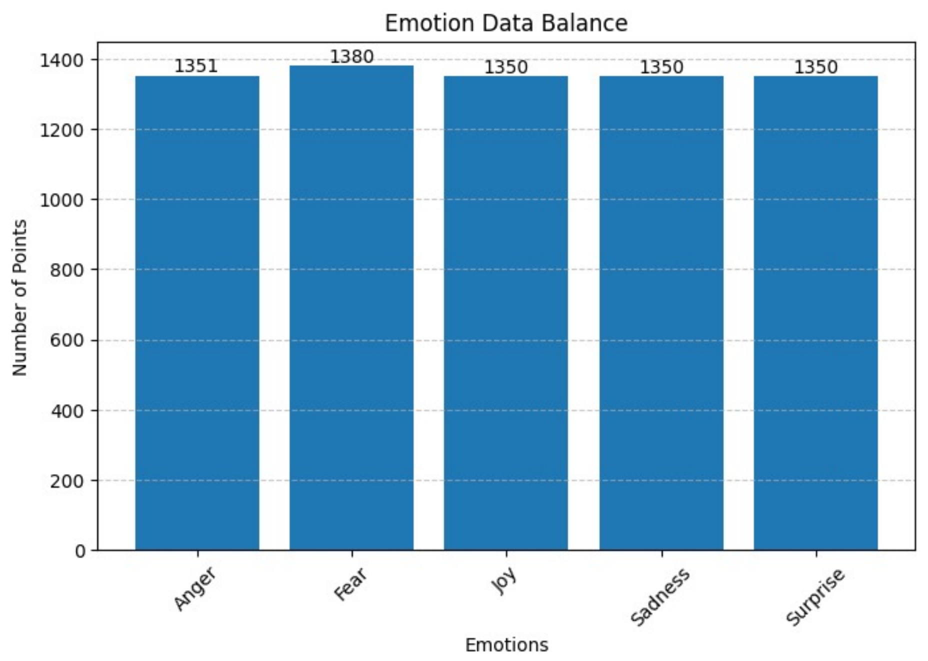
`Format: "text, Anger, Fear, Joy, Sadness, Surprise"`

`Labeling: Each label is either 1 or 0, indicating the presence or absence of that emotion in the text.`

`Task: Input Context: I have provided 100 sample entries representing texts labeled with emotions. Each entry reflects one or more emotions. These entries span a variety of tones, scenarios, and perspectives related to human experiences and reactions.`

`Objective: Using these 100 samples, generate 50 new synthetic entries. Ensure each new entry has similar language style, tone, and emotion combinations as the provided examples, maintaining the overall distribution and realistic variability.`

Data Distribution after Augmentation



Results after augmentation

Model Name	Accuracy	Micro F1	Macro F1	Weighted F1
SamLowe/roberta-base-go_emotions	0.40	0.59	0.48	0.55
cardiffnlp/twitter-roberta-large-emotion-latest	**0.61**	**0.76**	**0.74**	**0.72**
Emanuel/twitter-emotion-deberta-v3-base	0.24	0.49	0.44	0.45

Table 4: Results for Track A dataset after fine-tuning on the shortlisted transformer models after data augmentation

Future Directions

- Explore advanced data augmentation techniques like back-translation and generative methods.
- Investigate ensemble approaches to improve prediction accuracy.
- Develop custom architectures to better capture context and semantics.
- Expand cross-lingual emotion detection using multilingual models like XLM-Roberta.

Contribution

Task Done	Members Contribution
Literature survey for Text Based Emotion Detection	Equal contribution by all
Group discussions on literature survey	Equal contribution by all
Shortlisting the models based on literature survey	Equal contribution by all
Initial experimentation with the pretrained models	Equal contribution by all
Creation of presentation slides, Project Document and End-Term Project Report	Equal contribution by all

Table 5: Member Contribution

References

[1] Idris Abdulmumin. Semeval2025-task11, 2024. URL `https://github.com/emotion-analysis-project/SemEval2025-Task11`.

[2] Hassan Alhuzali and Sophia Ananiadou. Spanemo: Casting multi-label emotion classification as span-prediction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online and Punta Cana, Dominican Republic, 2021. Association for Computational Linguistics.

[3] Dimosthenis Antypas, Asahi Ushio, Francesco Barbieri, Leonardo Neves, Kiamehr Rezaee, Luis Espinosa-Anke, Jiaxin Pei, and Jose Camacho-Collados. Supertweeteval: A challenging, unified and heterogeneous benchmark for social media nlp research, 2023. URL `https://arxiv.org/abs/2310.14757`.

[4] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266, Marseille, France, June 2022. European Language Resources Association. URL `https://aclanthology.org/2022.lrec-1.27`.

[5] Sven Buechel and Udo Hahn. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, 2017.

[6] Ankush Chatterjee, Khyathi Raghavi Narahari, Meghana Joshi, and Puneet Agrawal. Semeval-2019 task 3: Emocontext contextual emotion detection in text. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 39–48, 2019.

[7] DAIR.AI. dair-ai/emotion, 2024. URL `https://huggingface.co/datasets/dair-ai/emotion/tree/main`.

[8] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions, 2020. URL `https://arxiv.org/abs/2005.00547`.

[9] Abhimanyu Dubey and et al. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

[10] Amir Gholipour Shahraki. Emotion mining from text, 2015. Available at: https://example.com.

[11] Pengcheng He, Jianfeng Gao, and Weizhu Chen. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2023. URL https://arxiv.org/abs/2111.09543.

[12] Emanuel Huber. Emanuel/twitter-emotion-deberta-v3-base, 2021. URL https://huggingface.co/Emanuel/twitter-emotion-deberta-v3-base.

[13] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019. URL https://arxiv.org/abs/1907.11692.

[14] Sam Lowe. Samlowe/roberta-base-go_emotions, 2022. URL https://huggingface.co/SamLowe/roberta-base-go_emotions.

[15] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, 2018.

[16] Emily Öhman, Marc Pàmies, Kaisla Kajava, and Jörg Tiedemann. XED: A multilingual dataset for sentiment analysis and emotion detection. In *The 28th International Conference on Computational Linguistics (COLING 2020)*, 2020.

[17] Daniel Preoțiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elizabeth Shulman. Modelling valence and arousal in facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15, 2016.

[18] Mahdi Rezapour. Emotion detection with transformers: A comparative study, 2024. URL https://arxiv.org/abs/2403.15454.

[19] Klaus R Scherer and Harald G Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2):310–328, 1994.