DEEP LEARNING (CS 583-C)

FINAL PROJECT

MENTAL HEALTH CHATBOT

Emotion Classification and Response Generation for Enhanced Mental Well-being

ARUN KASHYAP
Masters in Data Science
Stevens Institute of Technology
Fall 2024

PROBLEM STATEMENT

OBJECTIVE:

- Develop an AI-based chatbot to identify user emotions and provide empathetic responses. Enable mental health support in a scalable, accessible, and anonymous way.

USE CASE:

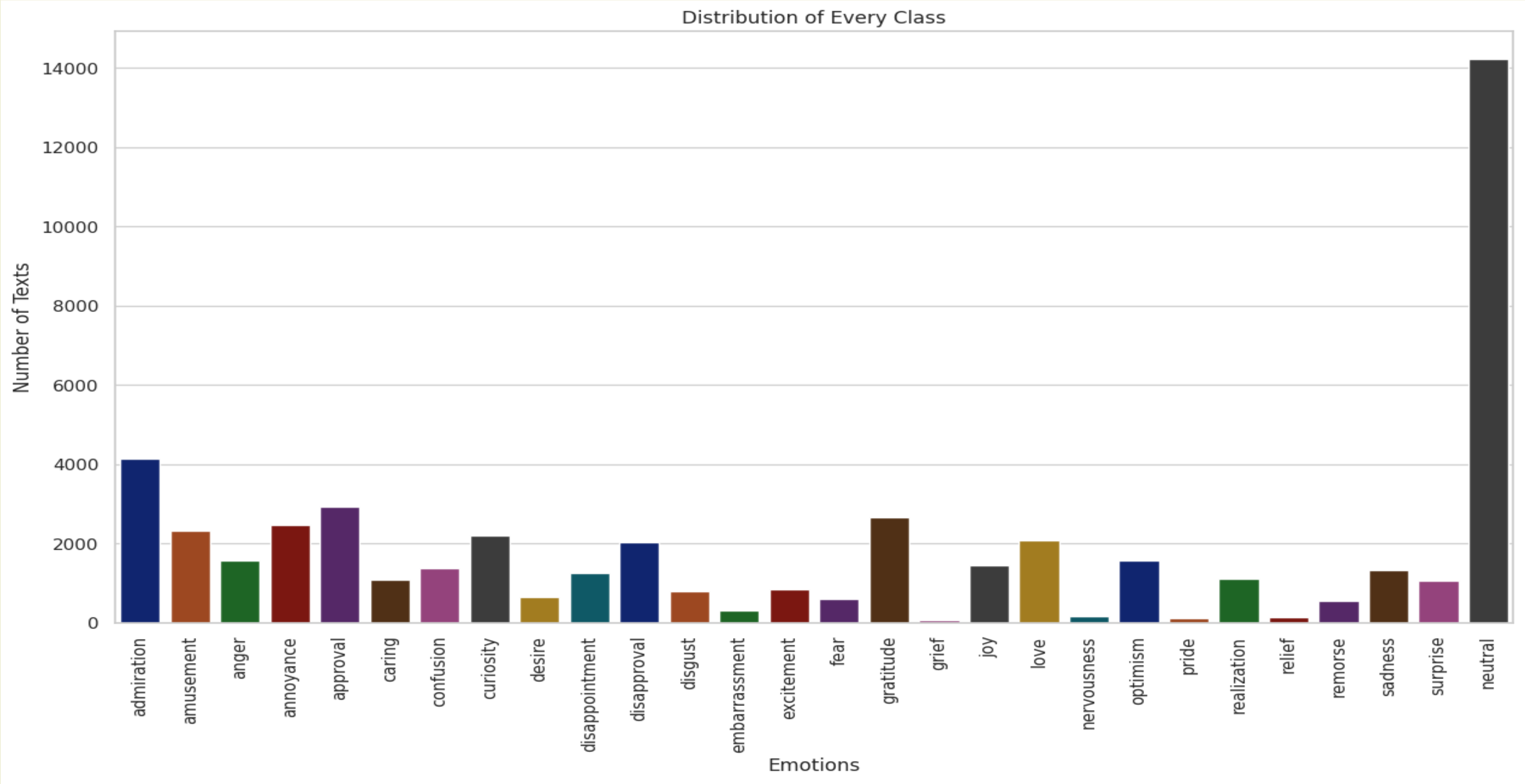
- Real-time emotional analysis for mental health self-care and conversational AI applications.

DATASET

- SOURCE:
  - The GoEmotions dataset by Google, consisting of 58k Reddit comments labeled with 28 fine-grained emotion categories.
- EMOTION CATEGORIES:
  - Consists of 28 emotions namely: "admiration", "amusement", "anger", "annoyance", "approval", "caring", "confusion", "curiosity", "desire", "disappointment", "disapproval", "disgust", "embarrassment", "excitement", "fear", "gratitude", "grief", "joy", "love", "nervousness", "optimism", "pride", "realization", "relief", "remorse", "sadness", "surprise", "neutral"
  - Primary emotions like "joy," "anger," "sadness," and "fear".
  - Includes Ekman's six basic emotions (anger, disgust, fear, joy, sadness, surprise).
- PREPROCESSING:
  - Tokenized with RoBERTa tokenizer.
  - Multi-hot encoded for multi-label classification.

EXPLORATORY DATA ANALYSIS

EMOTION DISTRIBUTION:



EXPLORATORY DATA ANALYSIS

EKMAN MAPPING:

JOY	ANGER	SADNESS	DISGUST	FEAR	SURPRISE
admiration	anger	disappointment	disgust	fear	surprise
amusement	annoyance	embarrassment		nervousness	realization
caring	disapproval	grief			confusion
excitement		remorse			curiosity
gratitude		sadness			
joy					
love					
optimism					
pride					
relief					

MODEL 1: BERT + LSTM

WHY DID I CHOOSE BERT + LSTM?

- Combines contextual embeddings (BERT) with sequential modeling (LSTM).
- Balances performance and computational efficiency for nuanced emotion classification.

Key Features:

- Lambda Layer
- Bidirectional LSTM: Enhances contextual understanding by processing token sequences both forward and backward.
- Dense Layers with Regularization:
  - ReLU activation reduces dimensionality.
  - Dropout (0.5) prevents overfitting.

MODEL 1: BERT + LSTM

MODEL COMPILATION AND OPTIMIZATION

- **Learning Rate Scheduler:**
  - *ReduceLROnPlateau*: Reduces learning rate dynamically to ensure stable convergence.
  - Initial learning rate: 1e-5.

- **Early Stopping:**
  - Monitors validation loss and stops training when no improvement is observed for 3 epochs, restoring the best weights.

- **Loss and Optimizer:**
  - Loss Function: Categorical Cross-Entropy for multi-class classification.
  - Optimizer: Adam optimizer ensures efficient gradient updates.

- **Regularization:**
  - Dropout layers with a rate of 0.5 minimize overfitting.

MODEL 1: BERT + LSTM

MODEL PERFORMANCE:

CLASSIFICATION REPORT:

- Low Overall Performance:
  - Accuracy: 19%Macro
  - Precision: 16%
  - Macro Recall: 21%
  - Macro F1-Score: 15%
  - Highlights challenges in differentiating emotions effectively.
- Class Imbalance and Misclassification:
  - Many low F1-scores, especially for minority emotions.
  - Near-zero precision and recall for several classes (e.g., labels 3, 12, 22).
  - Indicates inability to handle imbalanced data and overlapping emotional categories.

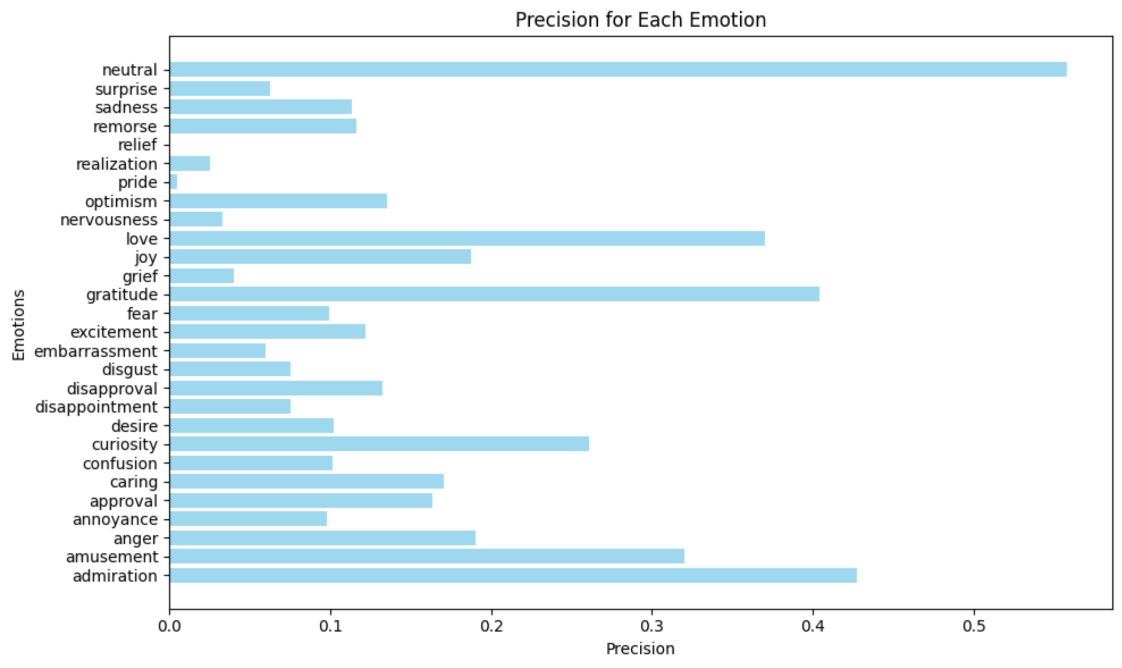
Classification Report:

	precision	recall	f1-score	support
0	0.43	0.11	0.17	863
1	0.32	0.25	0.28	427
2	0.19	0.29	0.23	321
3	0.10	0.06	0.08	411
4	0.16	0.04	0.06	523
5	0.17	0.49	0.25	191
6	0.10	0.15	0.12	237
7	0.26	0.64	0.37	368
8	0.10	0.28	0.15	109
9	0.08	0.02	0.03	191
10	0.13	0.37	0.19	311
11	0.08	0.29	0.12	117
12	0.06	0.06	0.06	49
13	0.12	0.30	0.17	149
14	0.10	0.16	0.12	101
15	0.40	0.70	0.51	423
16	0.04	0.15	0.06	13
17	0.19	0.04	0.07	218
18	0.37	0.24	0.29	319
19	0.03	0.09	0.05	23
20	0.14	0.16	0.15	190
21	0.00	0.08	0.01	12
22	0.03	0.01	0.01	136
23	0.00	0.00	0.00	14
24	0.12	0.46	0.19	78
25	0.11	0.24	0.15	188
26	0.06	0.24	0.10	143
27	0.56	0.06	0.11	2557
accuracy			0.19	8682
macro avg	0.16	0.21	0.15	8682
weighted avg	0.32	0.19	0.17	8682

MODEL 1: BERT + LSTM

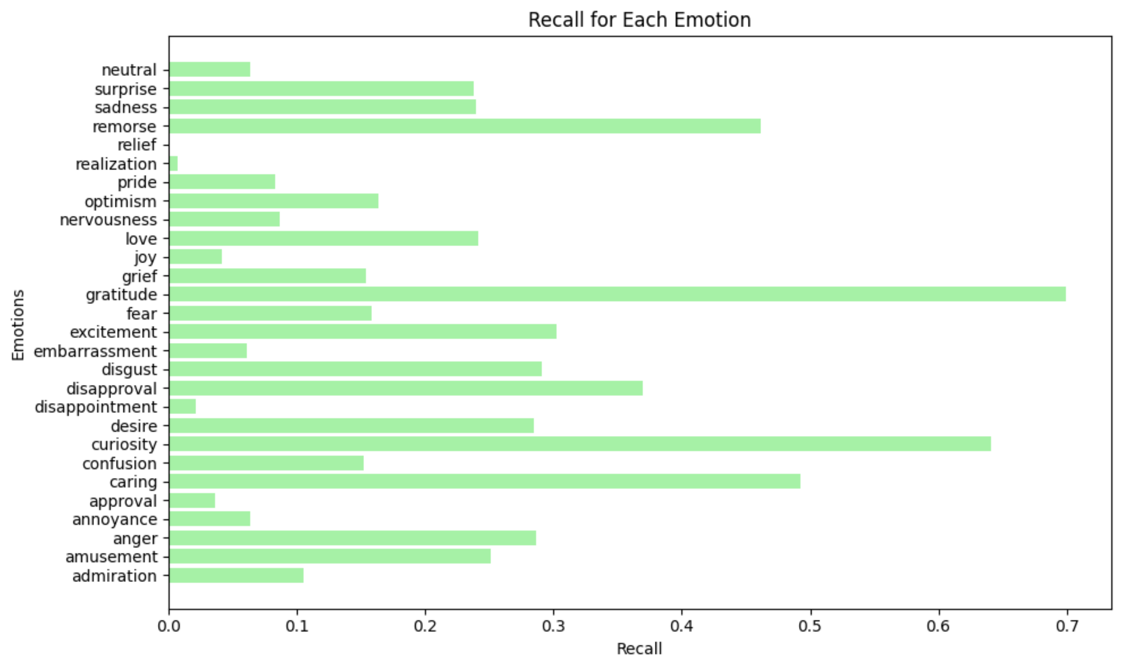
MODEL PERFORMANCE:

PRECISION



- Emotion Classification Gaps: The model struggles with subtle emotions like 'relief' and 'realization' while achieving better precision for 'neutral'.
- Imbalanced Performance: Precision varies significantly, showing bias towards frequent labels.

RECALL

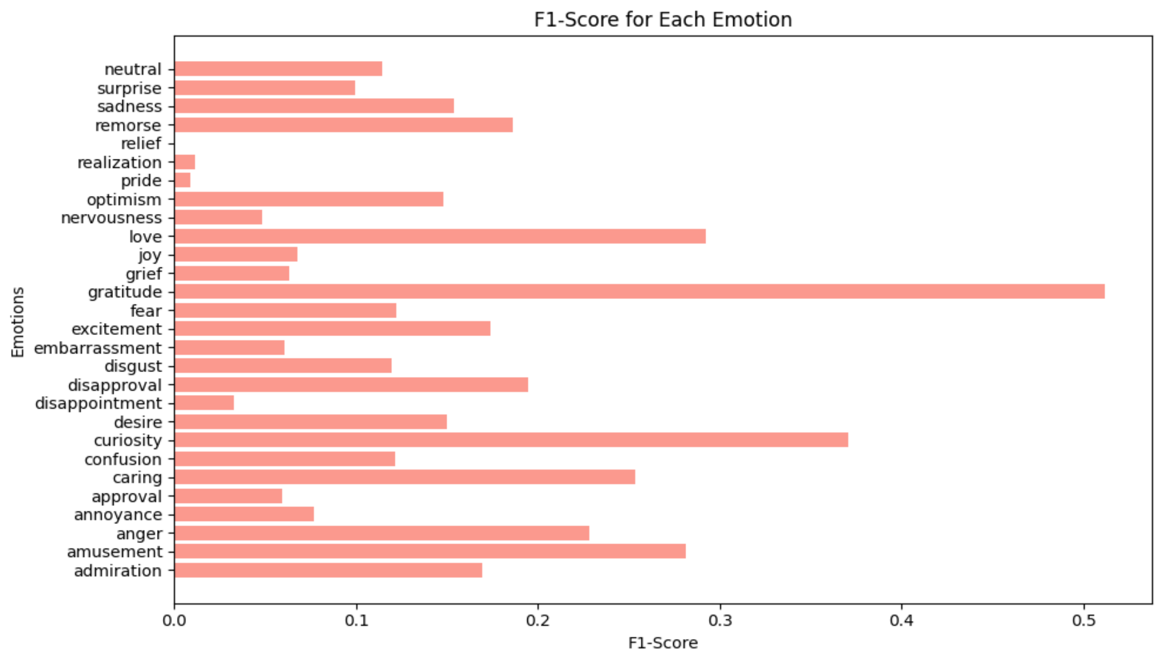


- Underrepresented Emotions: Emotions like "fear" and "joy" have higher recall, suggesting the model is good at detecting these labels but less so for "relief" and "realization.“
- Emotion Coverage: The disparity in recall values shows room for improvement in capturing less frequent emotions effectively.

MODEL 1: BERT + LSTM

MODEL PERFORMANCE:

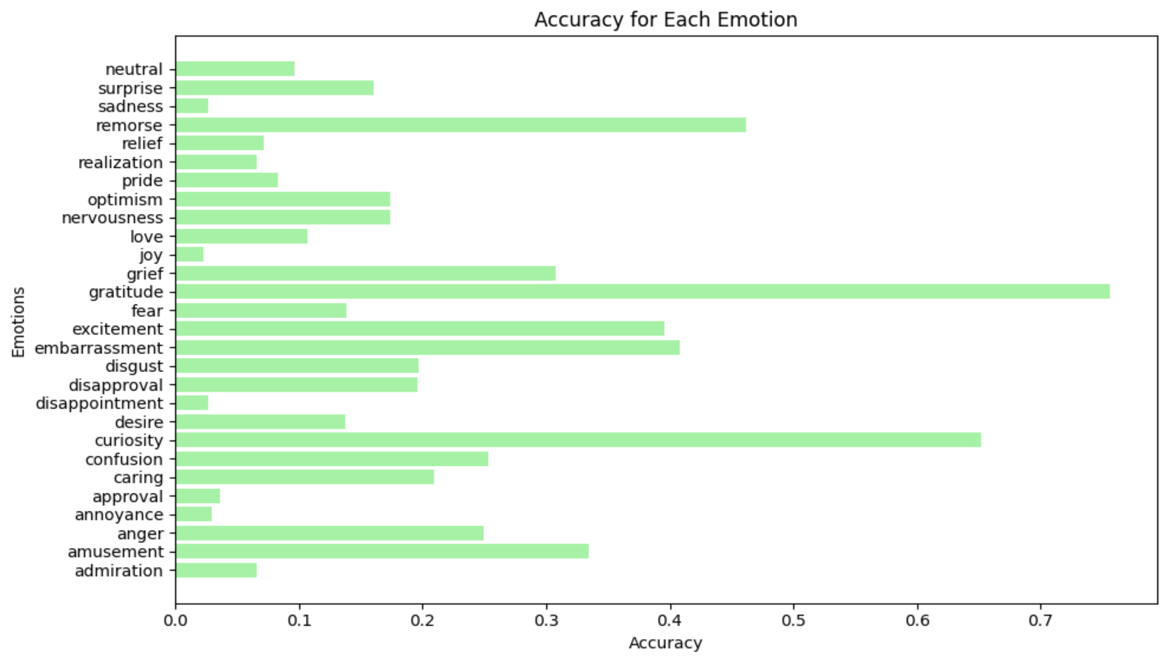
F1 SCORE



- Strengths: The model performs relatively well on emotions like "nervousness" and "curiosity," suggesting better handling of certain nuanced emotions.

- Weaknesses: Poor F1-scores for emotions like "relief" and "realization" highlight difficulties in managing underrepresented or context-dependent emotions.

ACCURACY



- Strengths: High accuracy for emotions like "neutral" and "joy" demonstrates the model's ability to classify common or well-defined emotional states effectively.

- Weaknesses: Low accuracy for emotions like "relief" and "pride" reflects challenges in addressing rare or overlapping emotional categories.

MODEL 2: RoBERTa (Robustly Optimized BERT Approach)

WHY DID I CHOOSE RoBERTa?

- State-of-the-art model for natural language understanding tasks.
- Optimized for fine-tuning tasks with larger datasets, like GoEmotions.
- Can handle multi-label classification better than BERT due to its advanced contextual embedding capabilities.

KEY FEATURES:

- Utilized RoBERTa as the Pretrained Language Model.
- Handled multiple emotion labels simultaneously with multi-hot encoding.
- Used contextual embeddings for deeper sentiment analysis.
- Leveraging Cosine Learning Rate Scheduler, Ensuring smoother and stabilized training convergence.
- Incorporates regularization (Dropout Layers) to prevent overfitting during training.

MODEL 2: RoBERTa

MODEL COMPILATION AND OPTIMIZATION

- **Optimizer:**
  - Implemented AdamW optimizer with a learning rate of 3e-5 for effective weight updates.

- **Learning Rate Scheduler:**
  - Used a Cosine learning rate scheduler to stabilize learning during training.
  - **Cosine Annealing:** Expect a smooth oscillation or gradual decay over time, depending on the scheduler.
  - **Learning Rate Drop:** For schedulers like ***ReduceLROnPlateau***, the plot will show sharp drops whenever the validation loss stagnates.

- **Loss Function:**
  - Adopted Binary Cross-Entropy (BCE) loss for precise multi-label emotion classification.

- **Batch Size and Epochs:**
  - Trained with a batch size of 32 for 7 epochs to ensure efficient processing and generalization.

MODEL 2: RoBERTa

MODEL PERFORMANCE:

CLASSIFICATION REPORT:

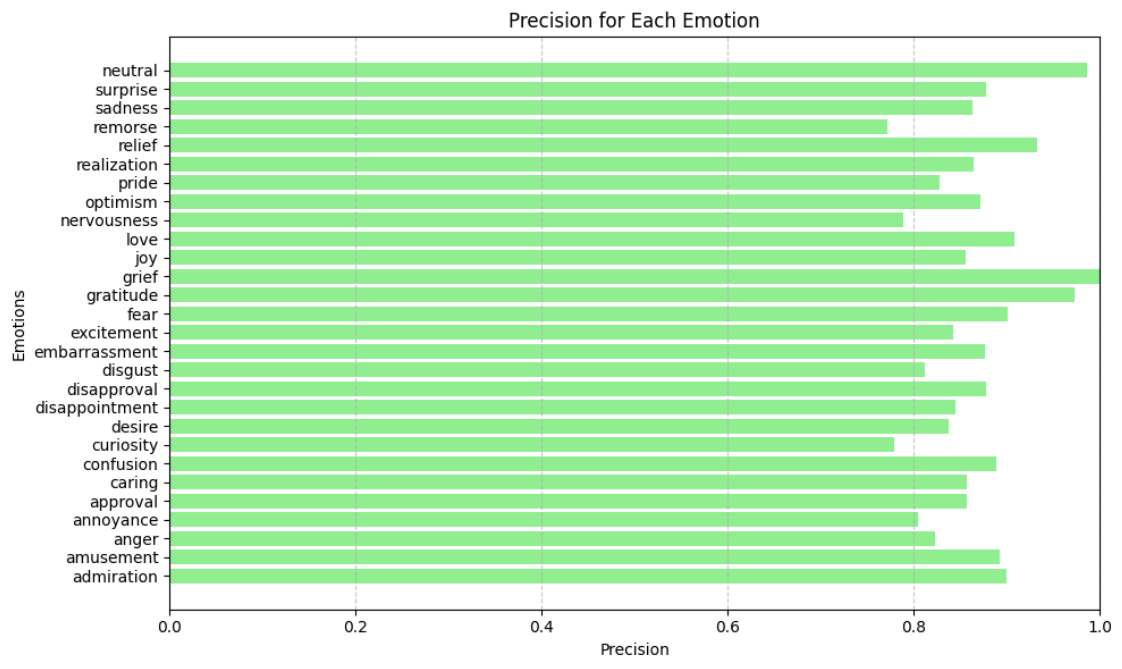
- Overall Performance:
  - Accuracy: 83%
  - Macro Precision: 87%
  - Macro Recall: 68%
  - Macro F1-Score: 74%
  - Demonstrates strong capabilities in distinguishing emotions effectively.
- Balanced Class Performance:
  - High F1-scores for majority classes (e.g., "joy," "gratitude," "anger"), indicating robust handling of frequent emotions..
  - Handles imbalanced data better, with improved precision and recall for minority emotions (e.g., "relief" and "realization"), though room for slight improvement remains.

	precision	recall	f1-score	support
Label 0	0.90	0.91	0.91	4130
Label 1	0.89	0.94	0.92	2328
Label 2	0.82	0.80	0.81	1567
Label 3	0.80	0.56	0.66	2470
Label 4	0.86	0.71	0.78	2939
Label 5	0.86	0.80	0.83	1087
Label 6	0.89	0.61	0.72	1368
Label 7	0.78	0.79	0.78	2191
Label 8	0.84	0.71	0.77	641
Label 9	0.84	0.54	0.66	1269
Label 10	0.88	0.76	0.81	2022
Label 11	0.81	0.65	0.72	793
Label 12	0.88	0.63	0.74	303
Label 13	0.84	0.66	0.74	853
Label 14	0.90	0.89	0.90	596
Label 15	0.97	0.93	0.95	2662
Label 16	1.00	0.01	0.03	77
Label 17	0.86	0.73	0.79	1452
Label 18	0.91	0.95	0.93	2086
Label 19	0.79	0.52	0.63	164
Label 20	0.87	0.73	0.79	1581
Label 21	0.83	0.43	0.57	111
Label 22	0.86	0.51	0.64	1110
Label 23	0.93	0.09	0.17	153
Label 24	0.77	0.79	0.78	545
Label 25	0.86	0.80	0.83	1326
Label 26	0.88	0.85	0.86	1060
Label 27	0.99	0.81	0.89	14219
micro avg	0.90	0.78	0.84	51103
macro avg	0.87	0.68	0.74	51103
weighted avg	0.90	0.78	0.83	51103
samples avg	0.87	0.82	0.83	51103

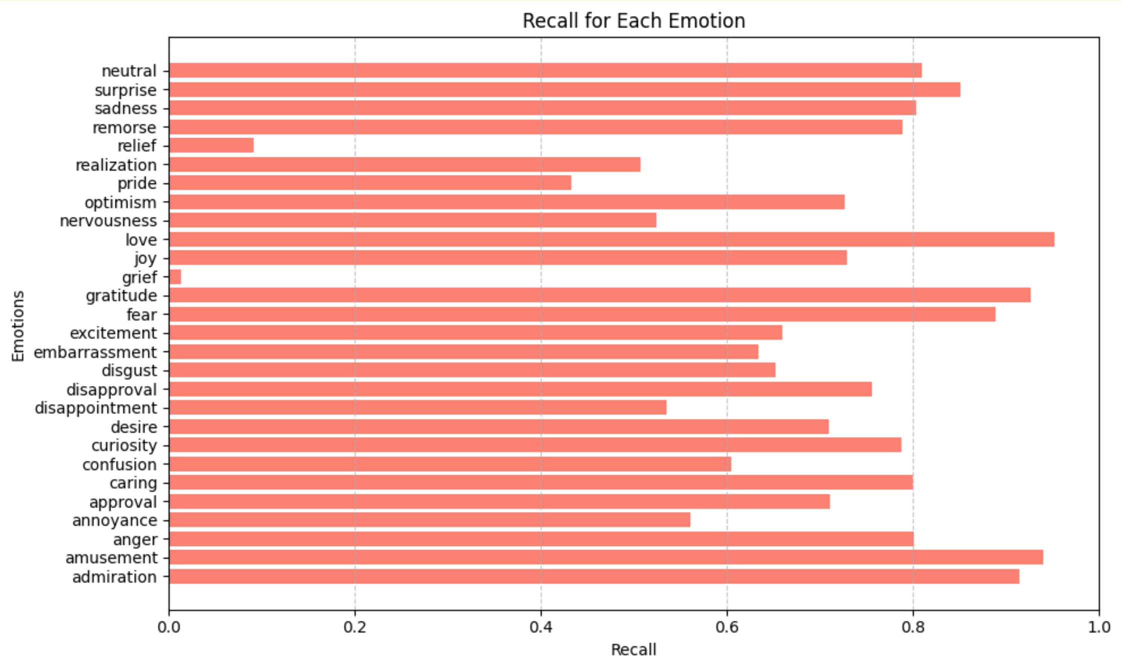
MODEL 2: RoBERTa

MODEL PERFORMANCE:

PRECISION



RECALL



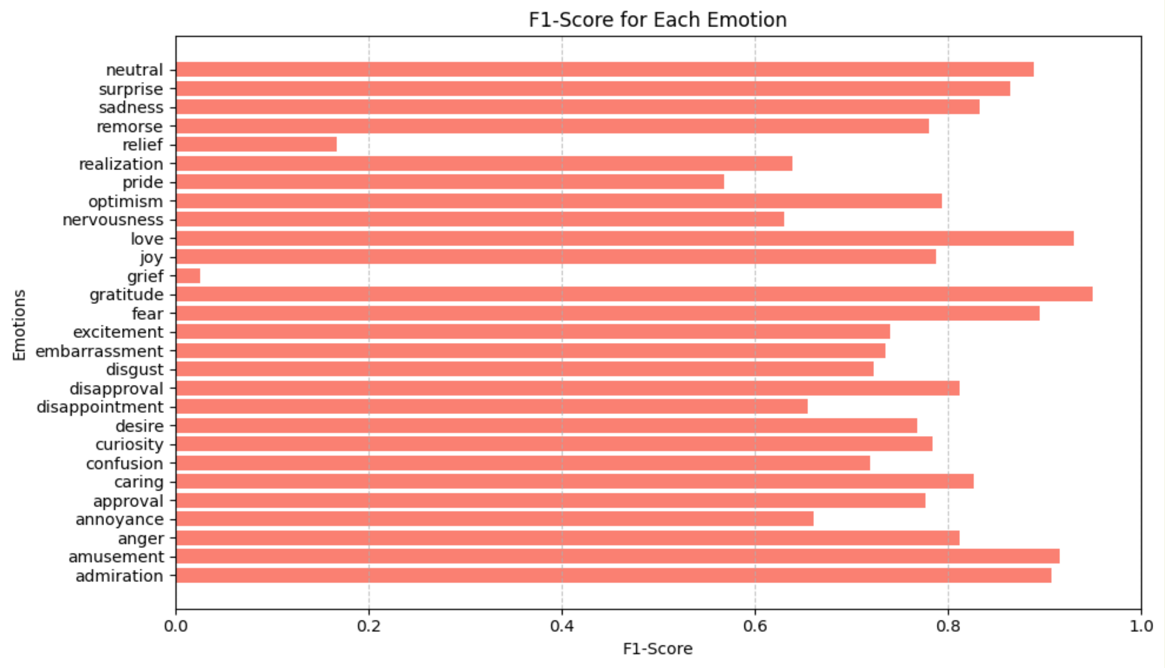
- Consistent Performance Across Emotions: It shows higher and more consistent precision values across emotions, better multi-label classification.
- High Precision for Common Emotions: Most of the emotions achieve precision values close to 0.80 and some 0.90, indicating the model's strength in distinguishing commonly occurring emotions.

- Strong Recall for Positive Emotions: Recall values for "joy," "admiration," and "gratitude" are significantly high, suggesting the model's effectiveness in correctly identifying these emotions.
- Moderate Recall for Difficult Emotions: While recall is generally high, some emotions like "fear" and "disgust" show moderate recall, reflecting slight challenges in capturing less frequent or nuanced emotional expressions.

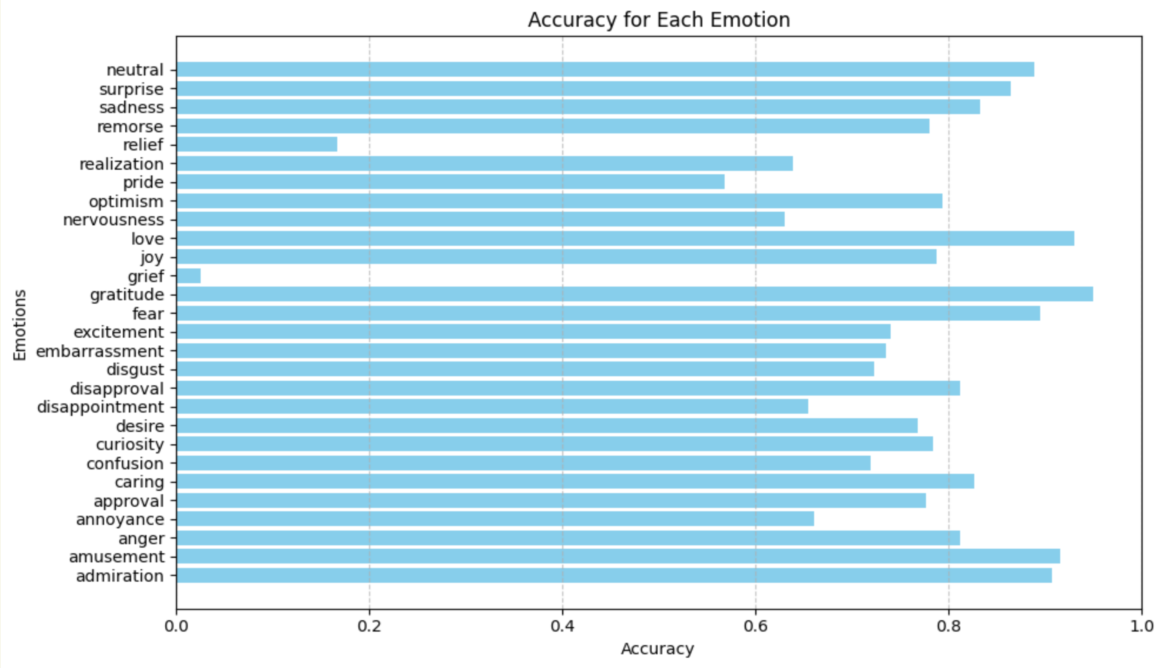
MODEL 2: RoBERTa

MODEL PERFORMANCE:

F1 SCORE



ACCURACY



- High F1-Scores for Common Emotions: Labels such as "neutral" and "joy" achieve higher F1-scores, reflecting balanced precision and recall for frequently occurring emotions.
- Variance in Performance for Rare Emotions: Lower F1-scores for emotions like "relief" and "realization" indicate the model's struggle with imbalanced data and complex representations.

- High Accuracy for Common Emotions: Emotions such as "neutral," "joy," and "admiration" exhibit high accuracy, reflecting the model's reliable predictions for frequent or well-represented classes.
- Lower Accuracy Complex Emotions: Emotions like "relief" and "grief" show lower accuracy, highlighting the model's difficulty in generalizing to less frequent or contextually nuanced emotions.

MODEL COMPARISON

No.	Criteria	BERT+LSTM	RoBERTa
1	Macro F1-Score	0.15	0.74
2	Weighted F1-Score	0.19	0.83
3	Handling Class Imbalance	Poor – Struggled with rare emotions	Good – Balanced performance across classes
4	Computational Efficiency	Computationally heavier (LSTM overhead)	Efficient due to streamlined architecture
5	Multi-Label Capability	Limited effectiveness	Optimized with binary cross-entropy loss
6	Precision & Recall	Low across most emotions	High and consistent across emotions
7	Training Stability	Prone to instability due to additional layers	Stable with cosine learning rate scheduler
8	Overall Recommendation	Not suitable for multi-label tasks	Ideal for robust emotion detection

CHATBOT INTERFACE

KEY FEATURES:

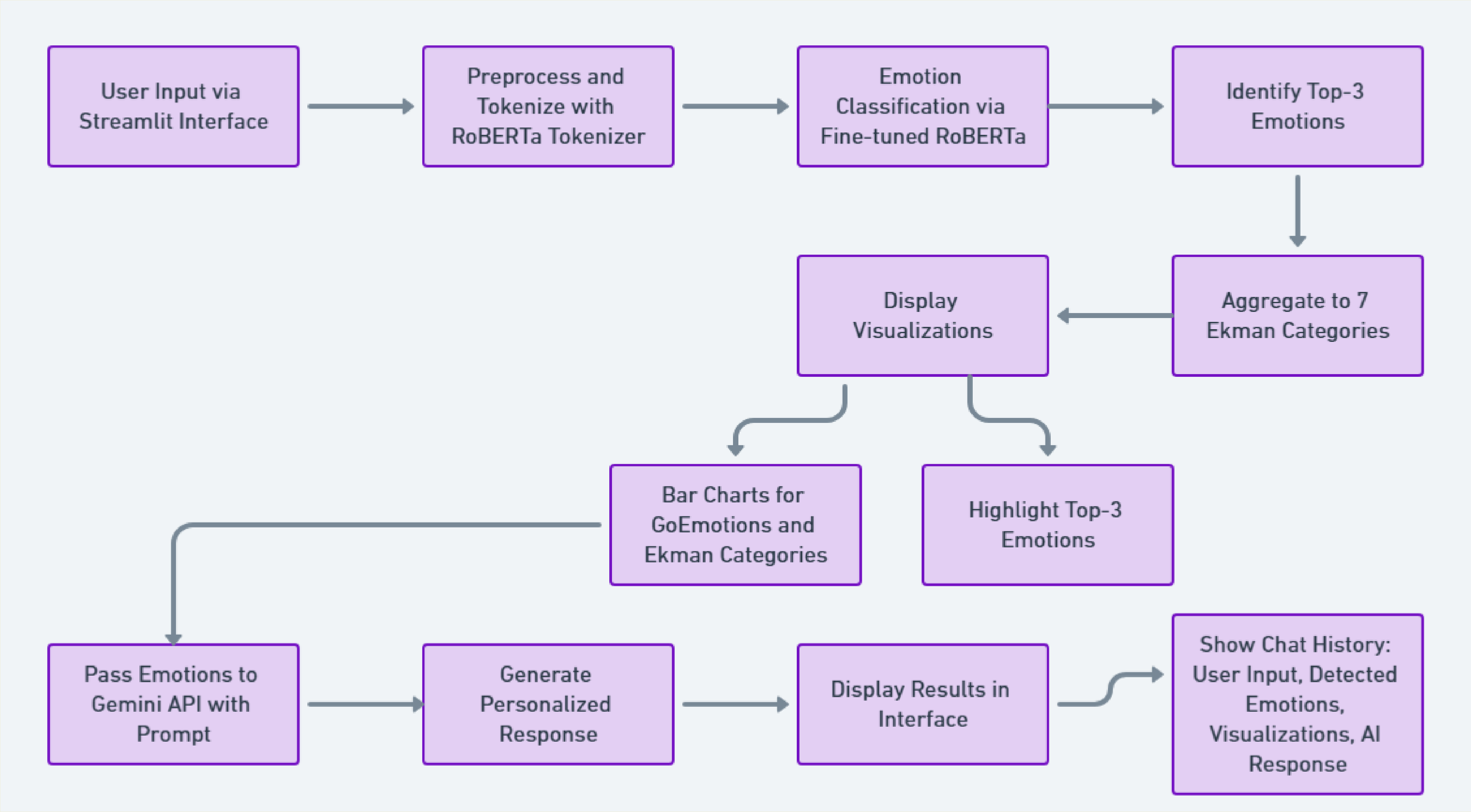
- Emotion Detection:
  - Identifies 28 emotions from the GoEmotions dataset.
  - Maps detected emotions to Ekman categories for better understanding.

- Custom Model:
  - Fine-tuned RoBERTa-based architecture with dense layers.
  - Achieves high precision, recall, and F1-scores across emotions.

- Interactive Interface:
  - Built using Streamlit for a user-friendly chatbot experience.
  - Visualizes GoEmotions and Ekman probabilities as bar charts. Displays Top-3 detected emotions for actionable insights.

- Solution-Oriented Responses:
  - Integrated Gemini API to generate mental health responses.
  - Tailored prompts ensure actionable solutions for detected emotions.

WORKFLOW


User Input via Streamlit Interface
Preprocess and Tokenize with RoBERTa Tokenizer
Emotion Classification via Fine-tuned RoBERTa
Identify Top-3 Emotions
Aggregate to 7 Ekman Categories
Display Visualizations
Bar Charts for GoEmotions and Ekman Categories
Highlight Top-3 Emotions
Pass Emotions to Gemini API with Prompt
Generate Personalized Response
Display Results in Interface
Show Chat History: User Input, Detected Emotions, Visualizations, AI Response

Q & A