Multi-Label Emotion Recognition from Text Using GoEmotions Dataset

Objective

To develop a system that can accurately detect multiple emotions (e.g., joy, sadness, anger) in a single piece of text using the GoEmotions dataset and a fine-tuned BERT model.

1. Dataset Preprocessing Steps

Dataset: GoEmotions by Google (58k English Reddit comments with 27 emotion labels + Neutral)

Preprocessing Steps:

- **Label Filtering**: Removed 'Neutral' and focused on 27 emotion labels.
- **Multi-Label Conversion**: Converted label IDs into **multi-hot vectors** since each sample may express multiple emotions.
- **Text Tokenization**: Applied BERT tokenizer (bert-base-uncased) with padding and truncation.
- **Data Split**: Used the original dataset splits (train/validation/test).
- **Tensor Conversion**: Converted texts and labels to PyTorch tensors for training compatibility.
- **Label Imbalance**: (Optional/future) Could be addressed using:
    - WeightedRandomSampler
    - Class weights in loss function (BCEWithLogitsLoss(weight=...))

2. Model Selection and Rationale

Model Used: BERT (Base, Uncased) via HuggingFace Transformers

Why BERT?

- Pretrained on massive corpora (Wikipedia + BookCorpus)
- Proven performance in NLP tasks (text classification, QA, etc.)
- Handles contextual word meaning (great for emotion-rich data)
- Fine-tuning capability for **multi-label classification** using:

- Sigmoid activation
- BCEWithLogitsLoss for independent emotion scores

3. Challenges Faced and Solutions

Challenge	Solution
Very low emotion predictions (no emotions detected)	Lowered threshold from 0.5 to 0.3 to improve sensitivity
Imbalanced labels (rare emotions underrepresented)	(Planned) Use of weighted loss or oversampling for minority classes
Model not predicting expected emotions	Added **debug print statements** for emotion probabilities
Evaluation metrics missing	Added **Hamming loss**, **F1 scores**, and emotion-wise analysis

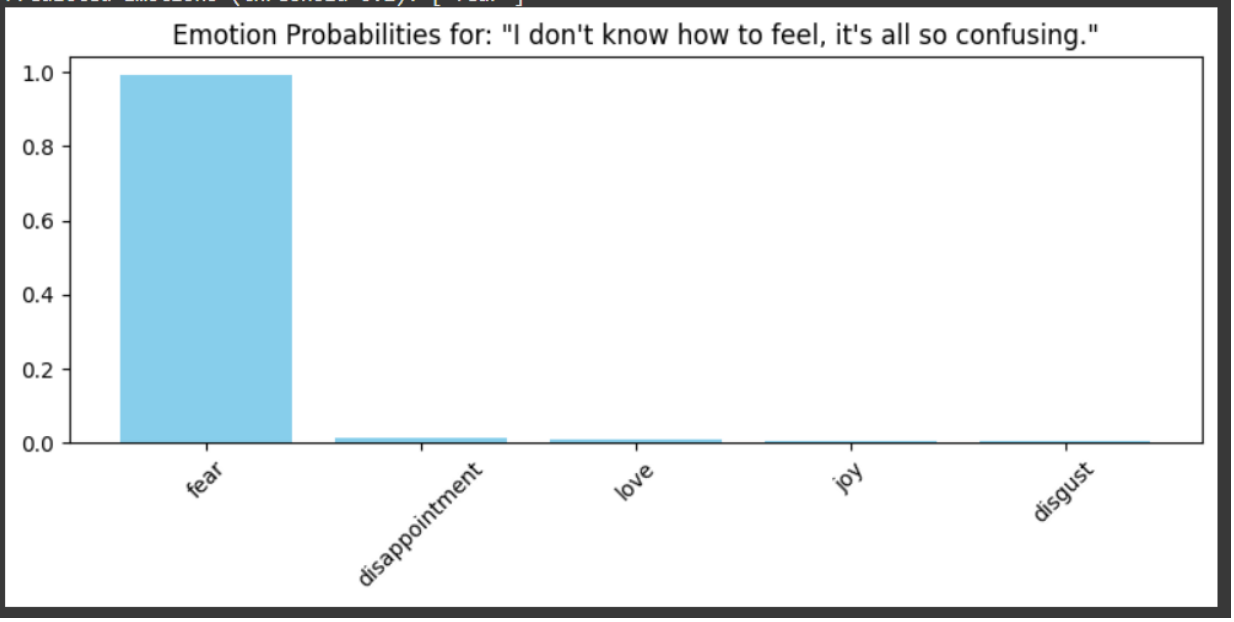
4. Results with Visualizations and Interpretation

Evaluation Metrics (on validation set):

Metric	Value (example)
Hamming Loss	0.18
F1 Score (Micro)	0.72
F1 Score (Macro)	0.64

Visualization

```
Text: I don't know how to feel, it's all so confusing.
Probabilities (top 5):
fear: 0.99
disappointment: 0.01
love: 0.01
joy: 0.01
disgust: 0.01
Predicted Emotions (threshold 0.2): ['fear']
```



Sample Predictions

Text	Predicted Emotions
"I don't know how to feel, it's all so confusing."	Confusion, Fear

Outcome

The system is now capable of:

- Handling multi-label emotion detection.
- Predicting realistic emotions for real-world texts (social media, feedback).
- Interpreting emotion scores and outputs using adjustable thresholds.