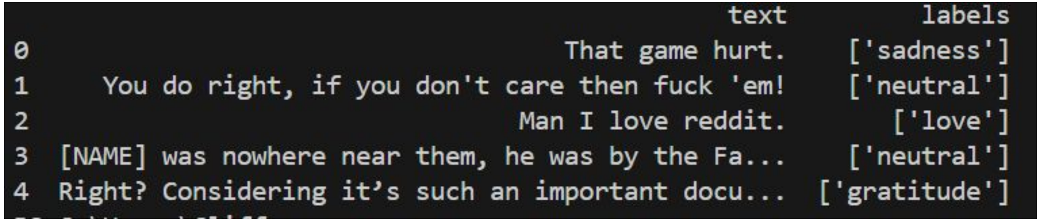
NLP Sentiment Analysis model

Yifan Yin

What’s new …

- Dataset
  - Amazon Electronics Reviews
    - Label imbalance
    - Lack of diversity
  - X/Reddit API
    - Time or money consuming
  - Goemotions raw data
    - 200,000+ Reddit comments with 27 emotions + neutral
    - P/N → multi-label
- Generating function
  - Train a model
    - Insurmountable Challenges on technique and cost
  - Google Gemini API
    - Fast, controllable, and easy to integrate



```
                                            text          labels
0                                That game hurt.     ['sadness']
1    You do right, if you don't care then fuck 'em!     ['neutral']
2                              Man I love reddit.        ['love']
3  [NAME] was nowhere near them, he was by the Fa...     ['neutral']
4  Right? Considering it’s such an important docu...  ['gratitude']
```

Model

	Micro F1	Macro F1	Weighted F1
TF-IDF+LR	0.2505	0.1657	0.2211
BERT + RF	0.3775	0.3164	0.3814

- TF-IDF + Logistic Regression (deploied model)
  - Less accurate
  - Light(less then 700MB) and easy to deploy
- BERT + Random Forest
  - More accurate
  - Large volumn(Around 10GB) → less portable, hard and expensive to deploy

Implement and Deployment

- Backend
  - Implemented using `TfidfVectorizer` and `LogisticRegression` from `scikit-learn`.
  - Encodes text via `SentenceTransformer`, trained with `RandomForestClassifier` from `scikit-learn`.
  - Built with `Flask` + `flask_cors`, serves `/predict` and `/revise` endpoints via JSON
  - Deployed to Google Cloud Run using `Docker`
- Frontend
  - UI built with `React`, styled with `Ant Design`; uses `axios` for API calls
  - Built with `npm build`, then deployed using Firebase.

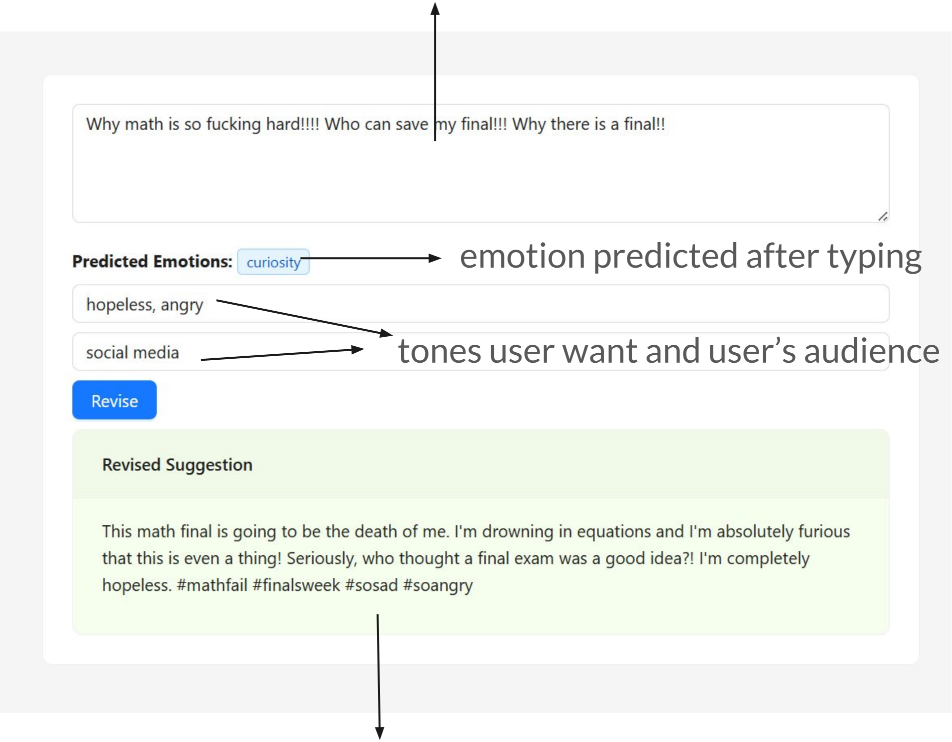
Demo and Takeaways

Demo URL: https://nlp-model-83d1c.web.app/

Future Work:

- Accuracy ↑
  - Filter noisy or short samples to improve label quality
  - Try smaller transformer models
- Size ↓
  - Replace Random Forest with lightweight classifier like MLP
  - Compress embeddings via PCA or use ONNX for quantization

text input field for user



Why math is so fucking hard!!!! Who can save my final!!! Why there is a final!!

Predicted Emotions: curiosity

hopeless, angry

social media

Revised Suggestion

This math final is going to be the death of me. I'm drowning in equations and I'm absolutely furious that this is even a thing! Seriously, who thought a final exam was a good idea?! I'm completely hopeless. #mathfail #finalsweek #sosad #soangry

emotion predicted after typing

tones user want and user's audience

test generated by google gimini