

From Words to Feelings, Emotion Prediction Using Classical and Deep Learning Models



Sohaib Bantan, Azzam Alharthi

Department of Computer Science and Eng,
University of Minnesota: Twin Cities, Minneapolis, MN, U.S.A

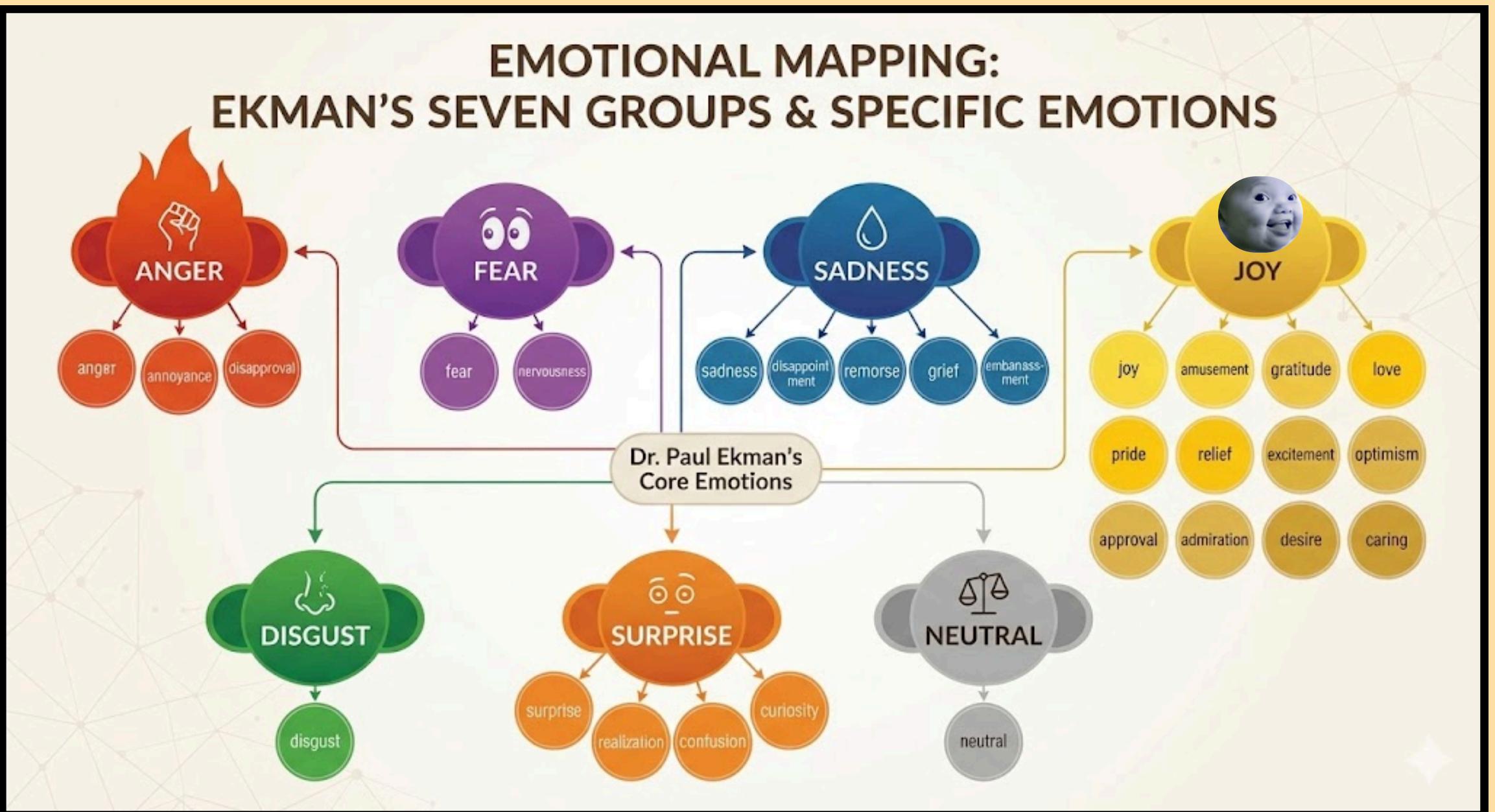
Introduction

MOTIVATION

- Emotion detection in text is crucial for mental health monitoring and sentiment mining.
- GoEmotions** is a dataset made up of Reddit comments that contain different emotions.
- We compare **classical** and **deep learning** models in classifying fine-grained emotions.

OBJECTIVES

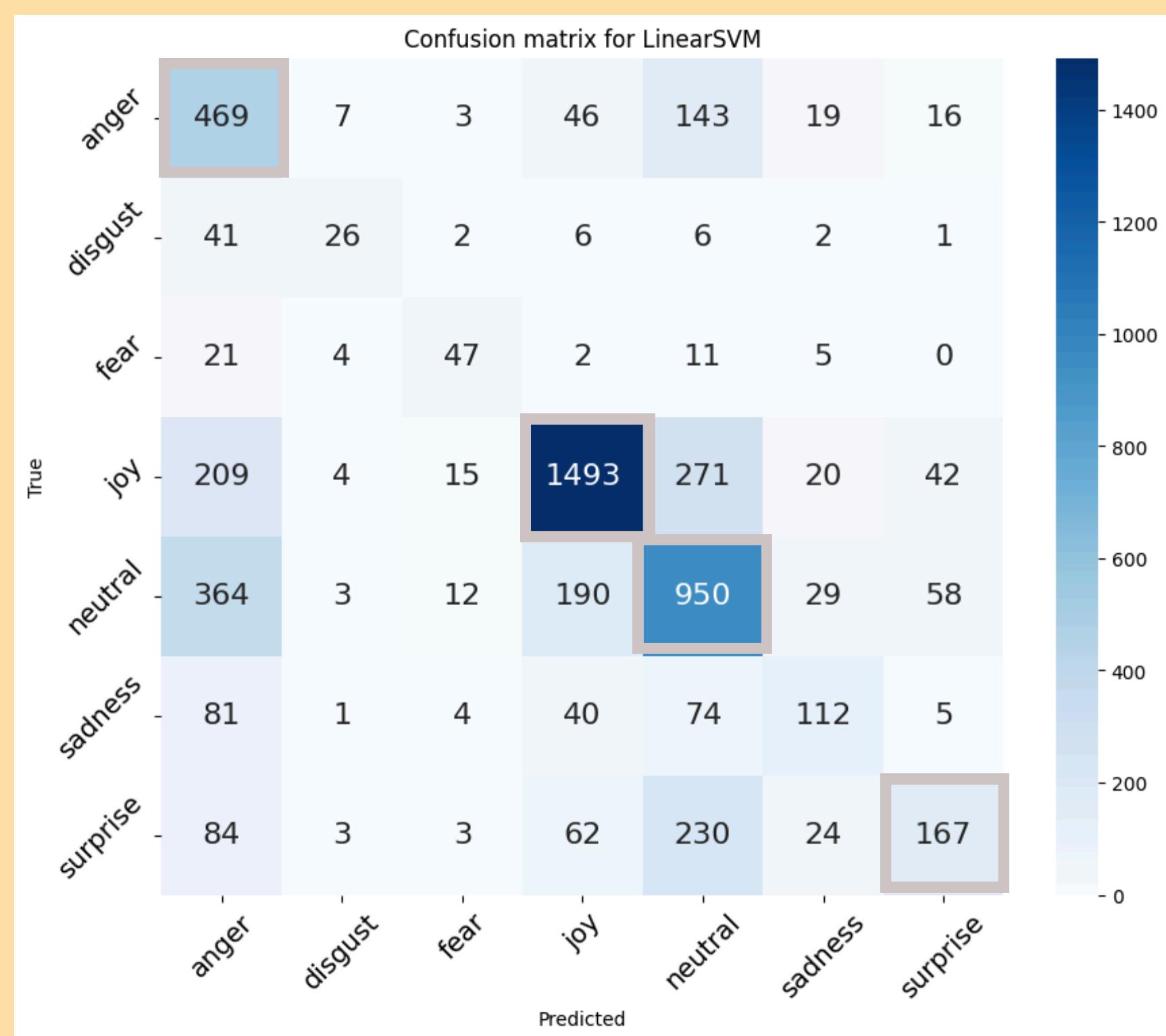
- Can machine learning and deep learning models accurately detect emotions from short text using Ekman's six emotion groups?
- How does a PyTorch neural network compare to a classical model and BERT-based model?
- Do these models generalize well to longer text inputs?



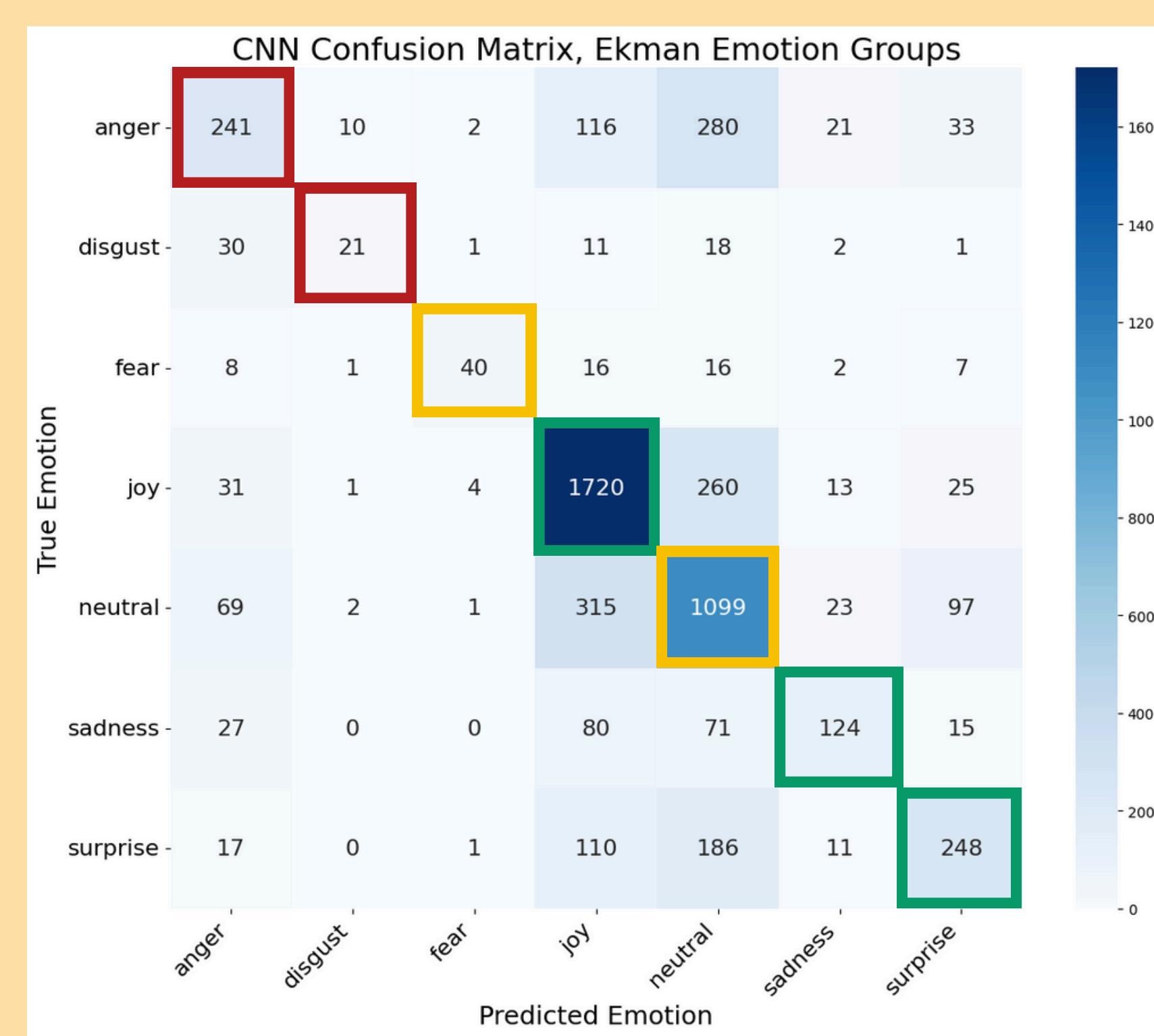
- GoEmotions, prepared by Google, consisted of **58,000** Reddit comments.
- Each comment is labeled with **28 emotions** plus neutral.
- Labels mapped into Paul Ekman's six core emotion groups.

Results & Discussion

LinearSVM MODEL



CNN MODEL



MACRO F1: 0.58

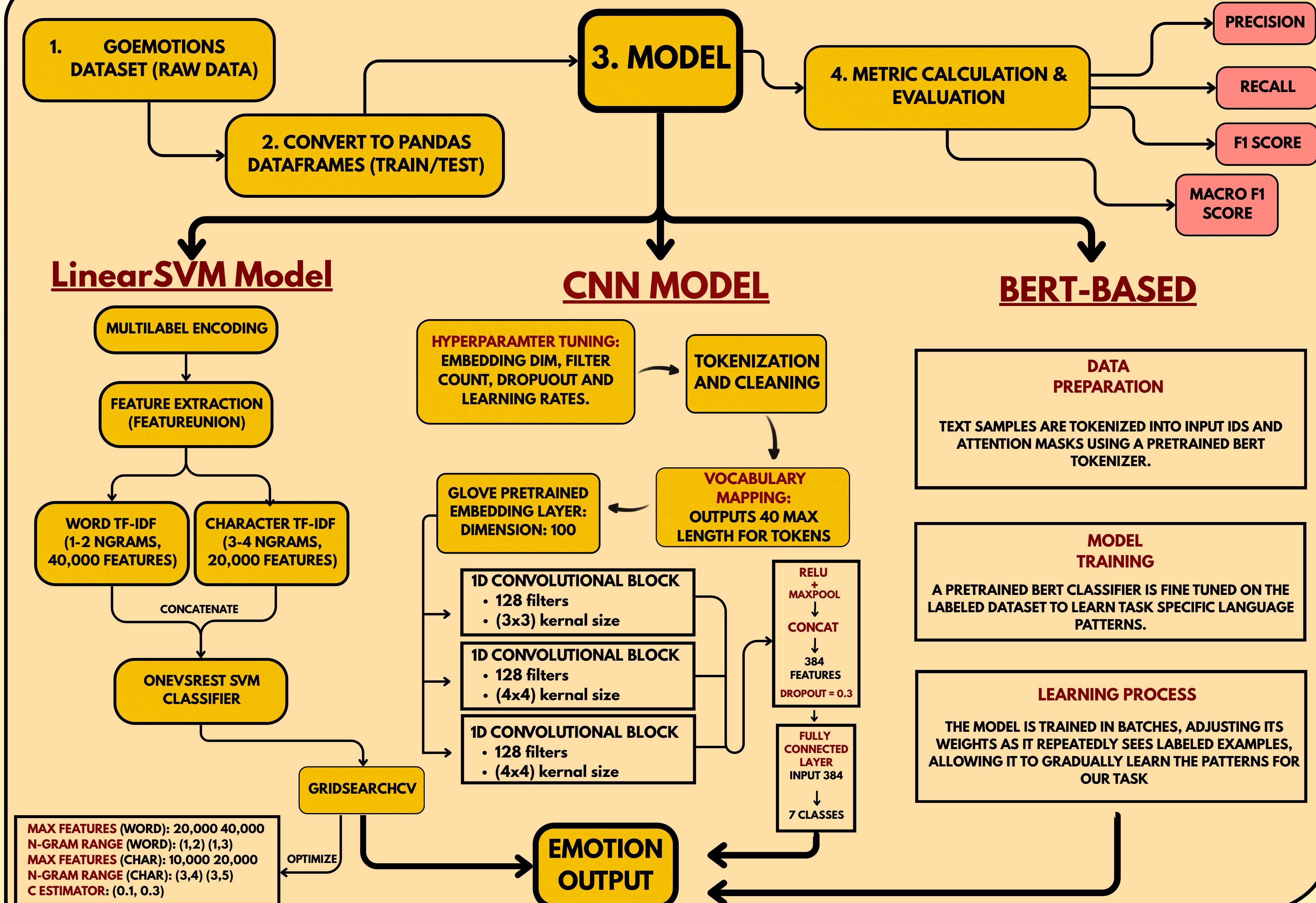
- Performed well for a classical model.
- Struggled with anger since comments contain sarcasm.

MACRO F1: 0.53

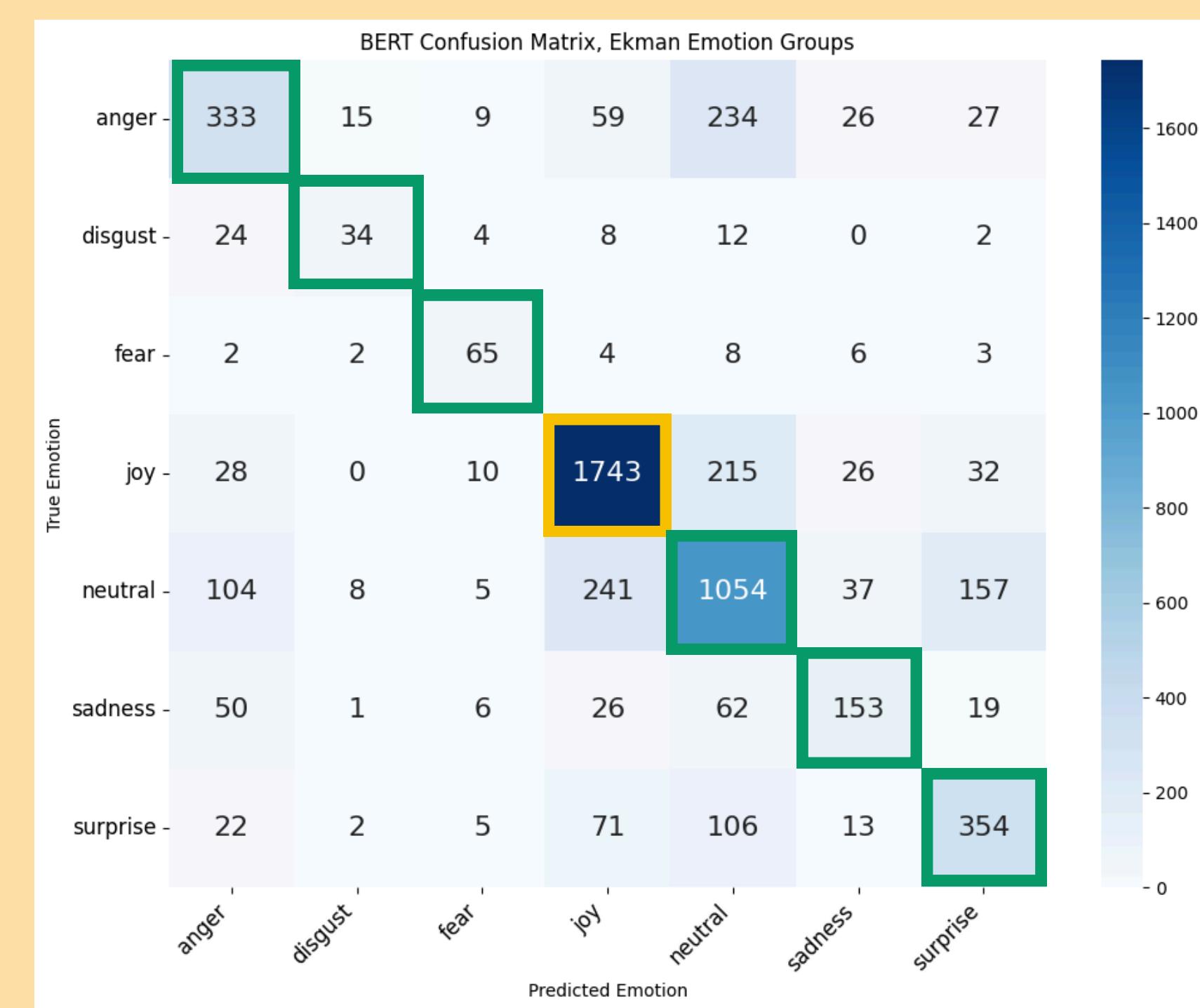
- More confusion in anger than SVM
- Captures surprise better

MACRO F1: 0.6 → Overall best score

Methodology



BERT-BASED MODEL



Conclusion

- BERT** gave the best results, which was expected from a pre-trained model.
- LinearSVM** gave great results because it used TF-IDF which performs best on short text samples.
- The **CNN** gave worse results than expected because it usually performs better on longer text samples instead of the short Reddit comments used in this project.
- The findings confirm that this task is feasible in the real world. That is because it greatly outperforms random guessing (about 4x better), which would have given a macro F1 of 0.14.

Acknowledgements

We express our gratitude to Dr. Bernardo Prado who provided us with guidance throughout this course and valuable feedback on this project. We also acknowledge Google Research for creating the high-quality GoEmotions dataset. We also thank the developers of Scikit learn and the pretrained BERT model.