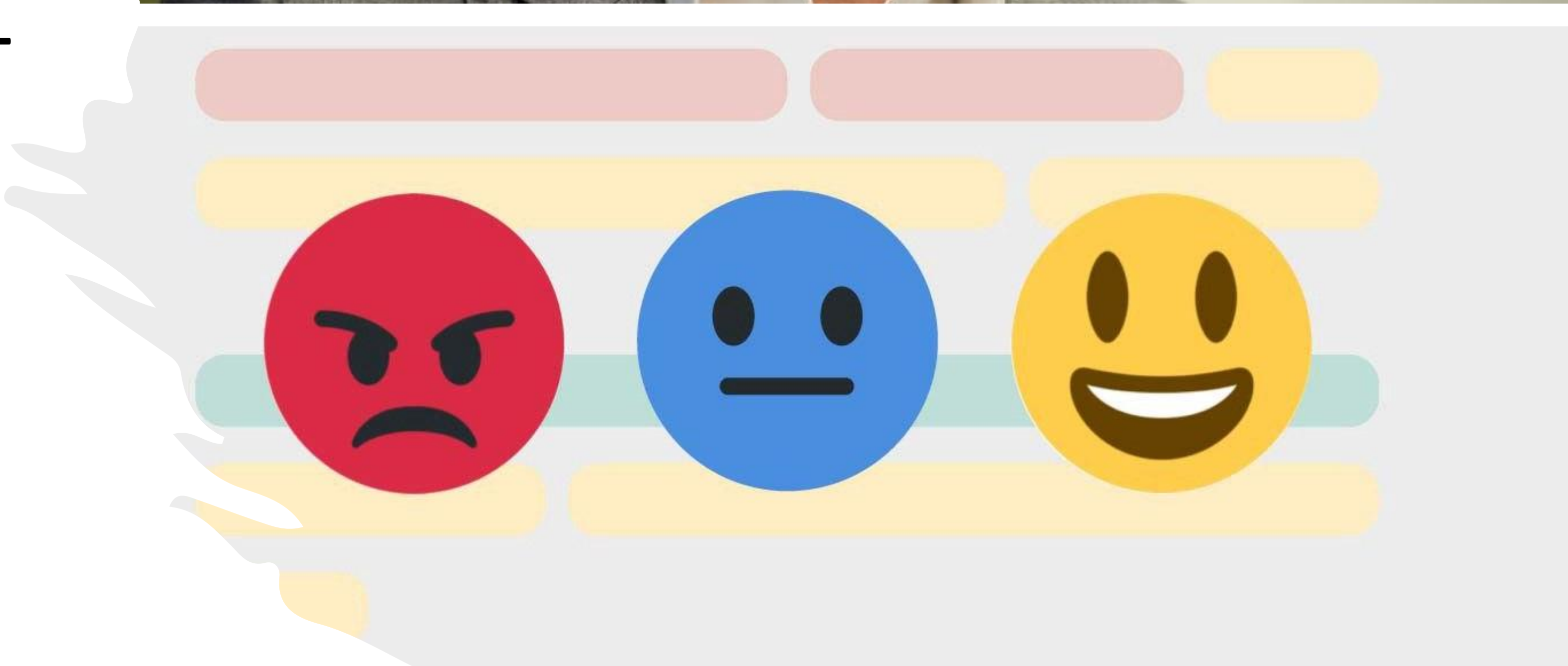
Emotion Classification using Tweets

Team ChatECT

aka. Good Luck Aim for DN+

- Cong Du (z5504217)
- Jiangye Song(z5270075)
- Jiayue Tang(z5471262)
- Pan Lu (z5530515)
- Xiangyu Zheng (z5465193)



Motivation

Goal

- Understanding emotions on social platform

Applications

- Disaster Response
- Review and Reactions
- Marketing
- Security and Threat Detection

Problem Statement

Problem

- Emotions are expressed in nuanced ways, varying by collective or individual experiences, knowledge, and beliefs.
- Understanding emotion through text requires a robust mechanism to capture and model different linguistic nuances and phenomena.

Aim

- To develop deep learning-based solutions to accurately identify emotions presented in textual data.

Presenter: Cong Du (z5504217)

Literature Review

"An LSTM model for Twitter Sentiment Analysis"

- **Author**: Md Parvez Mollah
- **Summary**:
- Utilizes an LSTM model for Twitter sentiment analysis.
- Combines seven annotated Twitter datasets for robust training and testing.
- Addresses overfitting with comprehensive preprocessing.
- Outperforms VADER model in identifying positive sentiments but requires more training time.

"FEEL-IT: Emotion and Sentiment Classification for the Italian Language"

- **Authors**: Federico Bianchi et al.
- **Summary**:
- Introduces FEEL-IT, a benchmark corpus for Italian sentiment and emotion classification.
- Annotates tweets with four emotions: anger, fear, joy, sadness.
- Uses BERT for context-based representation, achieving high performance.
- Provides an open-source Python library for further research.

Literature Review

"CARER: Contextualized Affect Representations for Emotion Recognition"

- **Authors**: Saravia, Elvis et al.
- **Summary**:
- Presents a semi-supervised, graph-based algorithm for deriving contextualized affect representations.
- Constructs datasets using distant supervision and enriches patterns with word embeddings.
- Outperforms state-of-the-art methods in emotion recognition tasks.
- Highlights potential applications in empathetic conversational agents.

"EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks"

- **Authors**: Muhammad Abdul-Mageed and Lyle Ungar
- **Summary**:
- Develops a dataset for fine-grained emotion detection using Twitter data and GRNNs.
- Models 24 types of emotions, achieving state-of-the-art performance.
- Uses distant supervision and validates data with human annotations.
- Extends emotion classification to Plutchik's eight basic emotions.

Dataset

Source

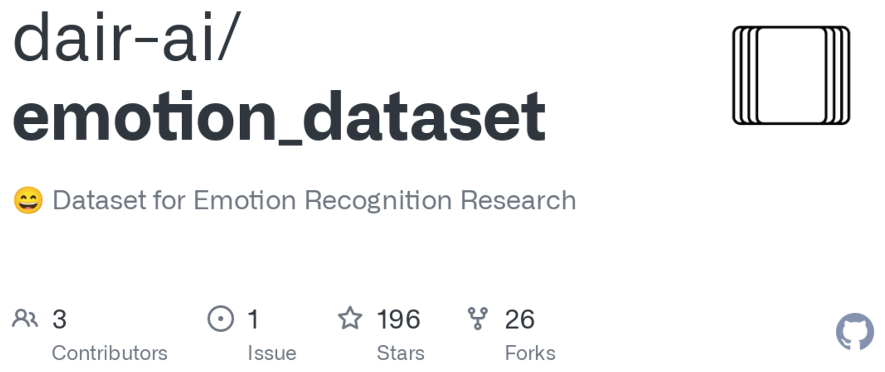
- EMOTION dataset from Hugging Face
- dair-ai/emotion (split)
- English Twitter messages

Composition

- Training set: 16,000 tweets
- Validation set: 2,000 tweets
- Test set: 2,000 tweets

Emotions

sadness, joy, love, anger, fear, surprise

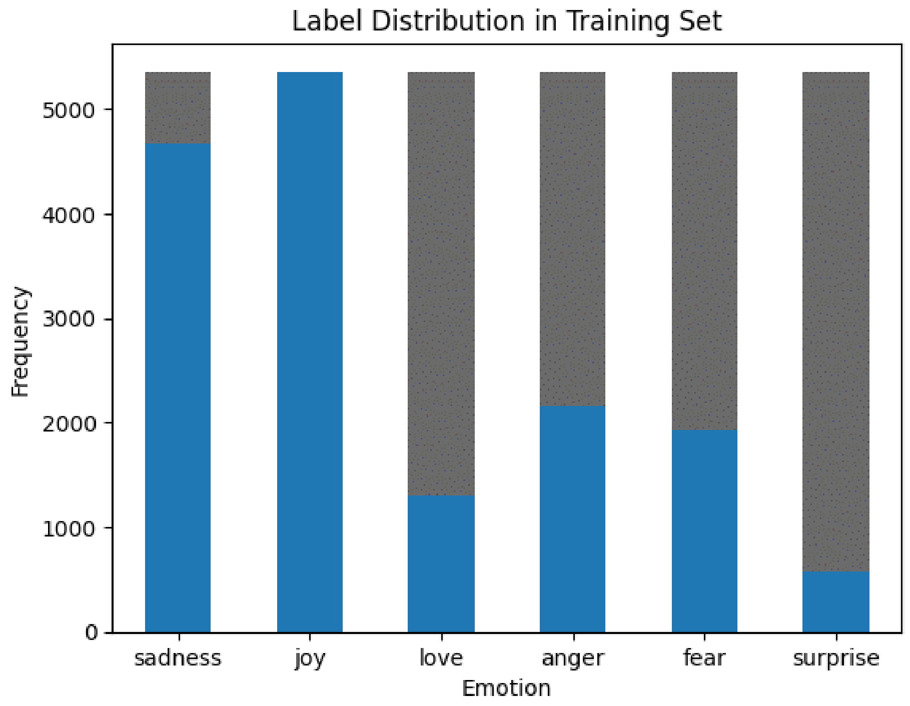


i didnt feel humiliated	0 sadness
i can go from feeling so hopeless to so damned hopeful just from being around someone who cares and is awake	0 sadness
im grabbing a minute to post i feel greedy wrong	3 anger
i am ever feeling nostalgic about the fireplace i will know that it is still on the property	2 love
i am feeling grouchy	3 anger
ive been feeling a little burdened lately wasnt sure why that was	0 sadness
ive been taking or milligrams or times recommended amount and ive fallen asleep a lot faster but i also feel like so funny	5 surprise

Presenter: Jiangye Song(z5270075)
Dataset URL: https://huggingface.co/datasets/dair-ai/emotion
Dataset paper: https://aclanthology.org/D18-1404/

Data Analysis

Label Distribution

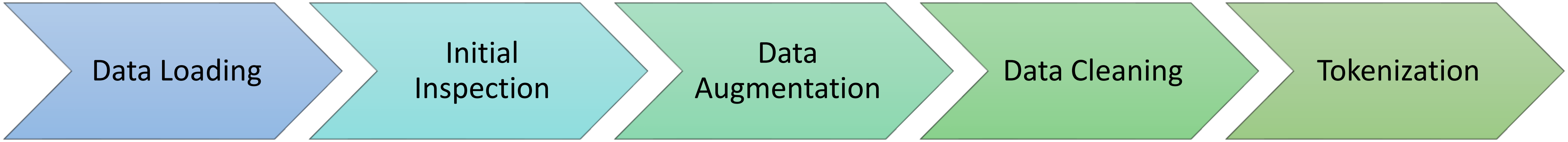


Data Augmentation

Presenter: Jiangye Song(z5270075)

Data Analysis

Data preprocessing



Presenter: Jiangye Song(z5270075)

Methods

Common Steps for all Models

- Trained on the same training, validation, and test sets
- Tokenize (Pre-processing)
- Use SoftMax as activation function
- Compiled with sparse categorical cross-entropy and default Adam optimizer
- Used early stopping to prevent overfitting

Long Short-Term Memory (LSTM) Model

Gated Recurrent Neural Network (GRU) Model

Convolutional Neural Network (CNN) Model

Methods

Common Steps for all Models

Long Short-Term Memory (LSTM) Model

- Embedding Layer (Dimension: 128)
- Spatial Dropout Layer
- Bidirectional LSTM Layer (64 units)
- Dense Layer

Gated Recurrent Neural Network (GRU) Model

Convolutional Neural Network (CNN) Model

Presenter: Xiangyu Zheng(z5465193)

Methods

Common Steps for all Models

Long Short-Term Memory (LSTM) Model

Gated Recurrent Neural Network (GRU) Model

- Embedding Layer (Dimension: 128)
- Spatial Dropout & Dropout Layer
- Two Bidirectional GRU Layers
- Dense Layer

Convolutional Neural Network (CNN) Model

Methods

Common Steps for all Models

Long Short-Term Memory (LSTM) Model

Gated Recurrent Neural Network (GRU) Model

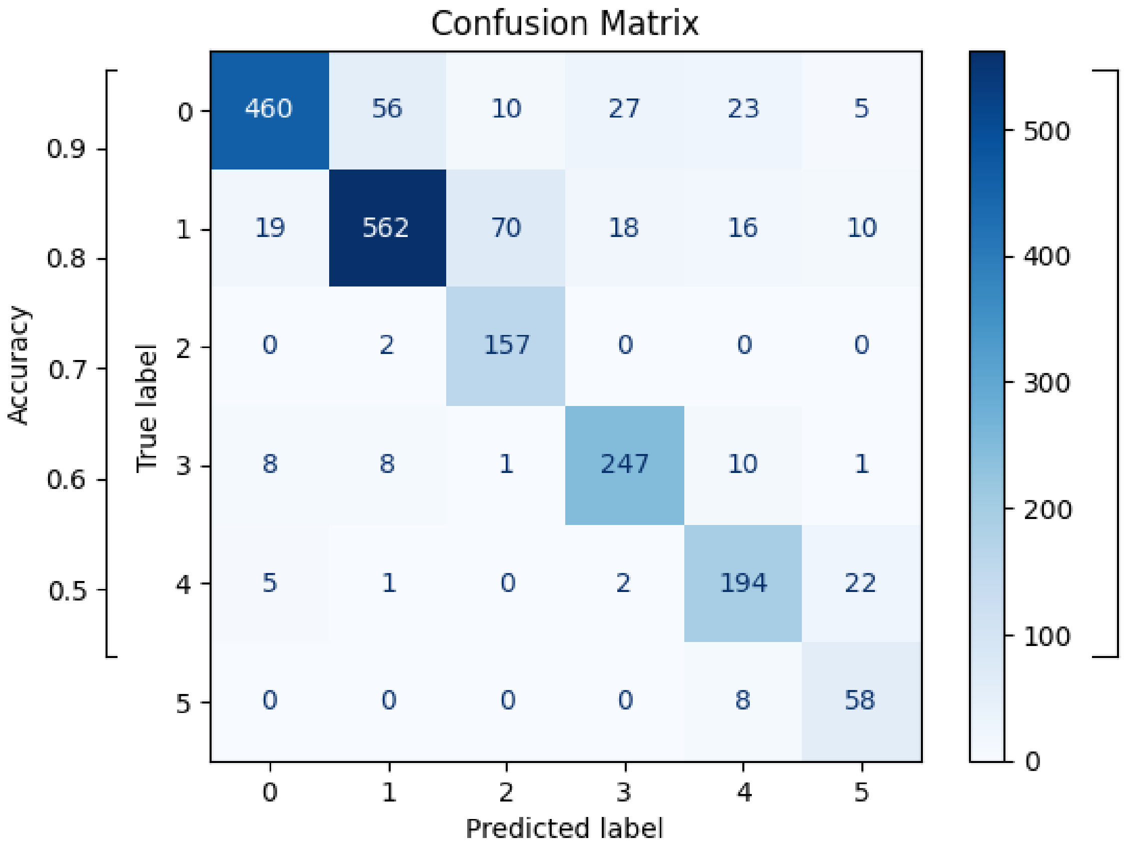
Convolutional Neural Network (CNN) Model

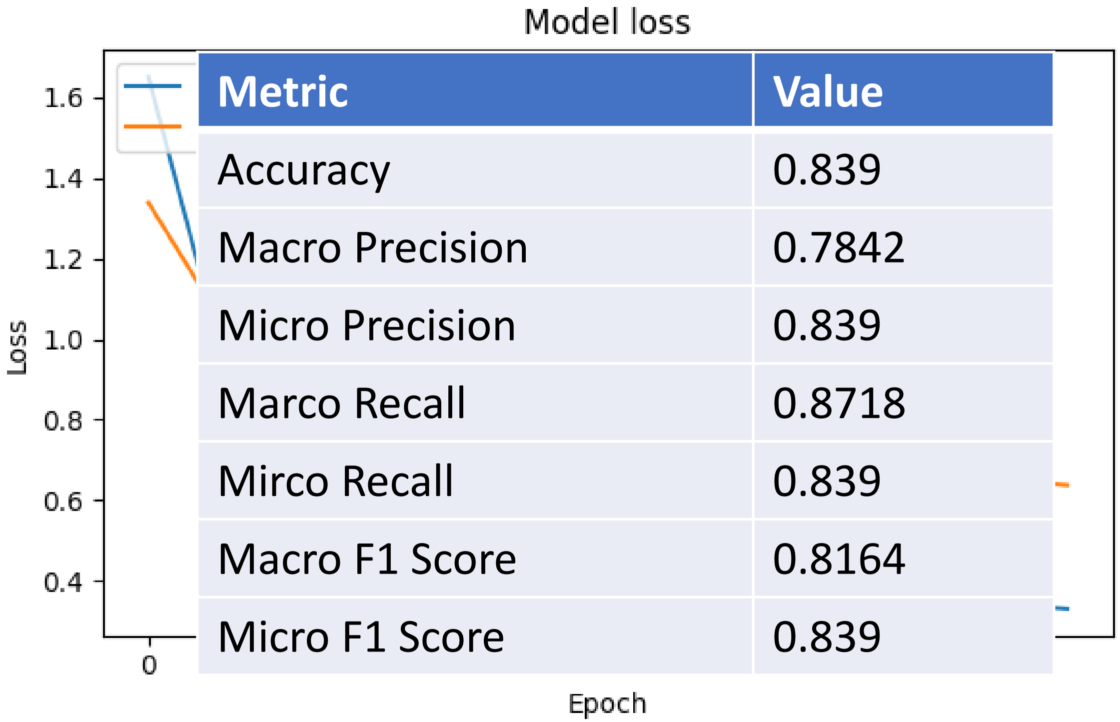
- Conv1D layers (128 & 64 filters)
- Embedding layer (Dimension: 100)
- Max Pooling Layer
- Flatten and Dense Layer
- Dropout Layer
- Dense Layer

Presenter: Xiangyu Zheng(z5465193)

Results

- LSTM

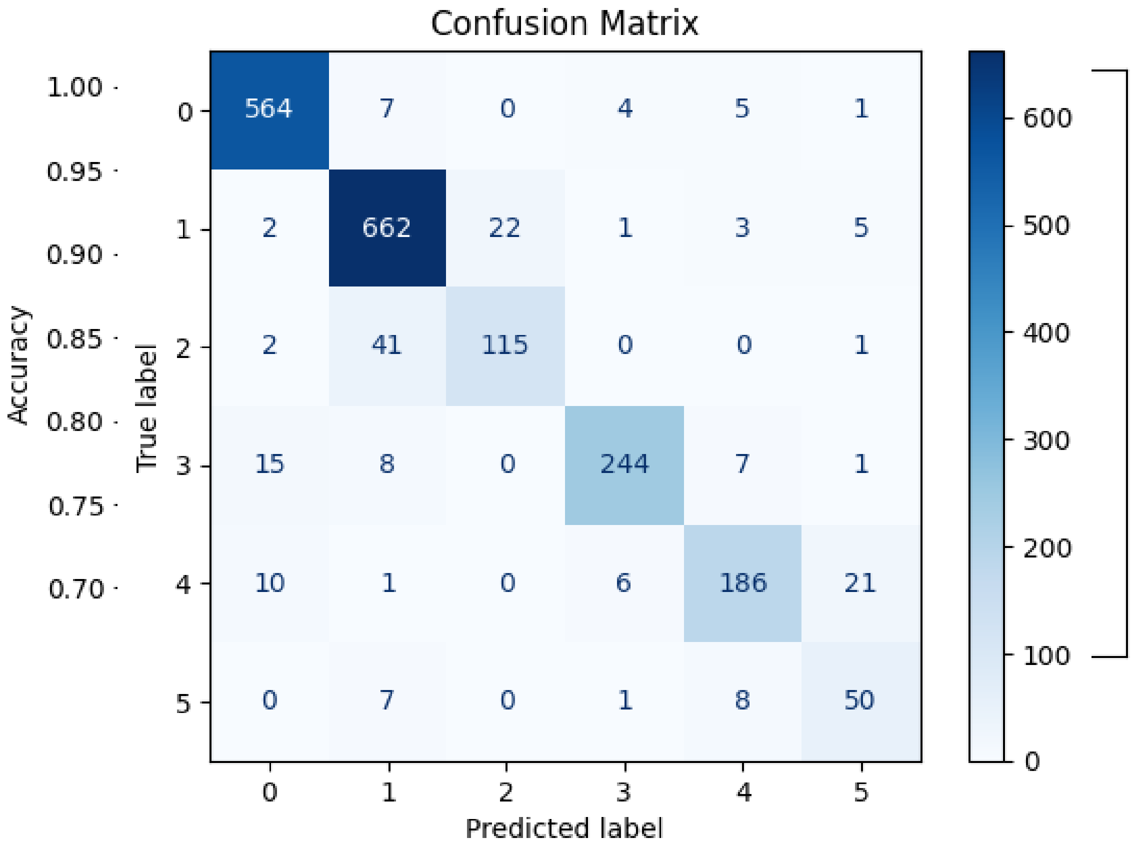


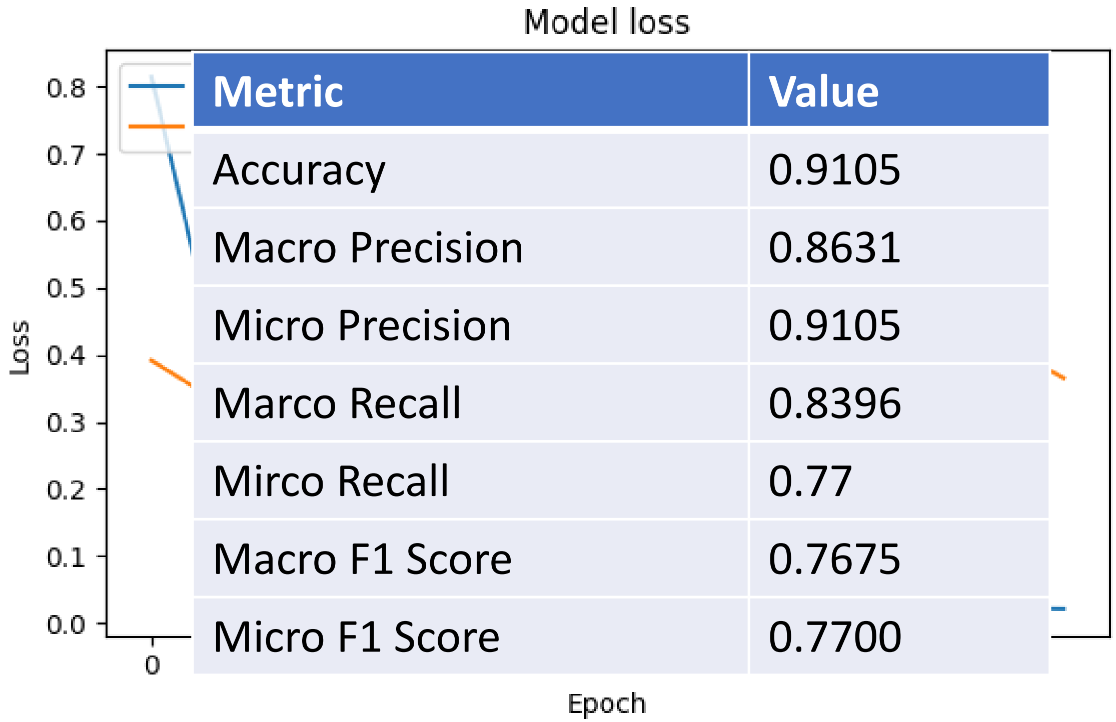


Metric	Value
Accuracy	0.839
Macro Precision	0.7842
Micro Precision	0.839
Marco Recall	0.8718
Mirco Recall	0.839
Macro F1 Score	0.8164
Micro F1 Score	0.839

Results

- GRU



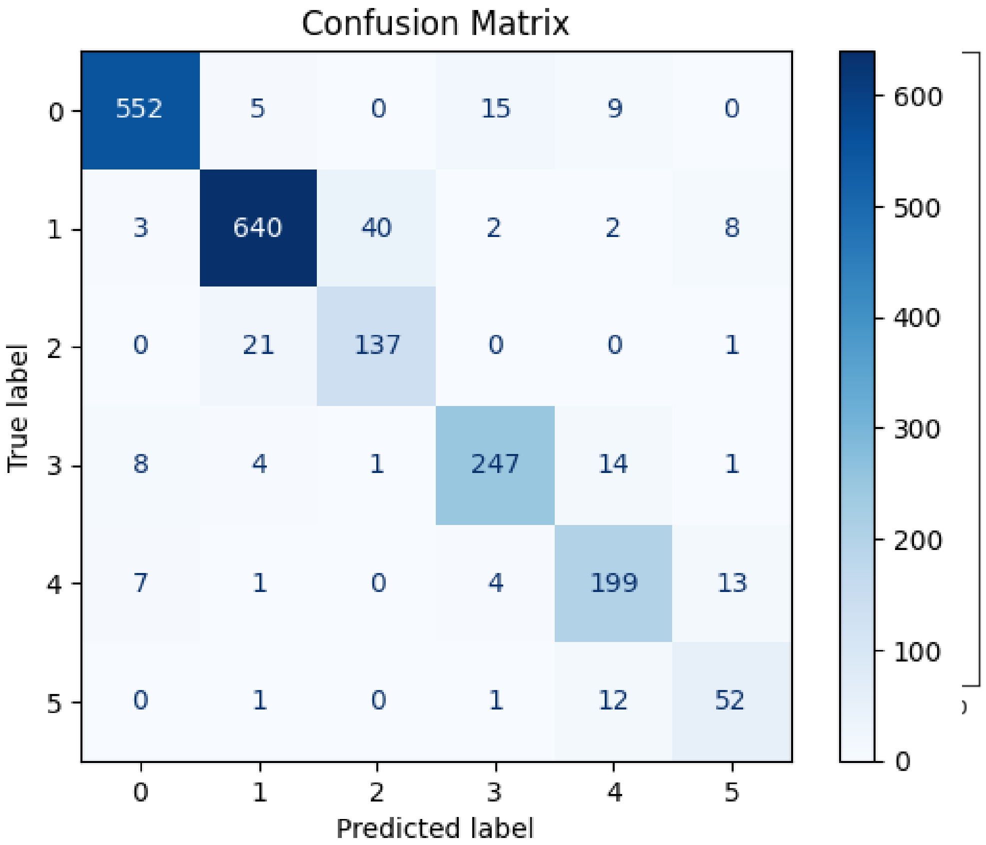


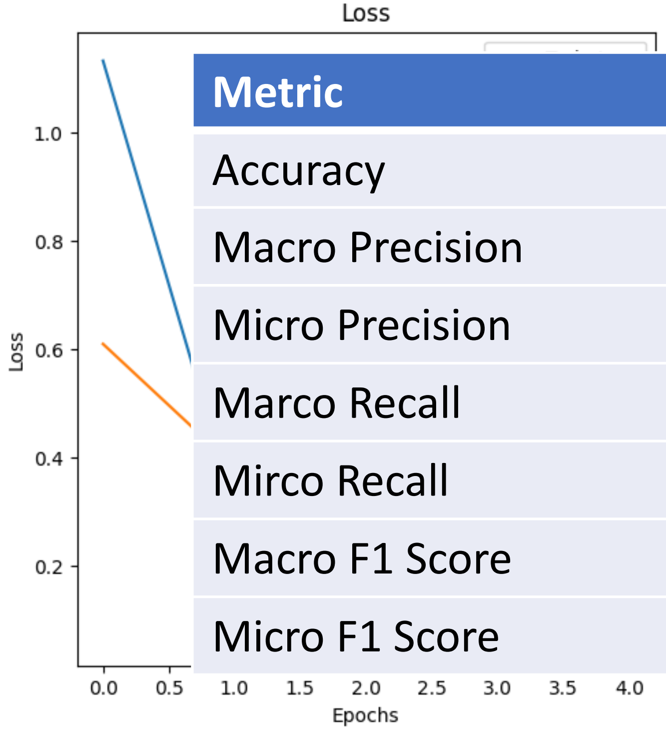
Metric	Value
Accuracy	0.9105
Macro Precision	0.8631
Micro Precision	0.9105
Marco Recall	0.8396
Mirco Recall	0.77
Macro F1 Score	0.7675
Micro F1 Score	0.7700

Presenter: Xiangyu Zheng(z5465193)

Results

- CNN





Metric	Value
Accuracy	0.9135
Macro Precision	0.8575
Micro Precision	0.9135
Marco Recall	0.8845
Mirco Recall	0.9135
Macro F1 Score	0.8699
Micro F1 Score	0.9135

Presenter: Xiangyu Zheng(z5465193)

Discussion

Long Short-Term Memory (LSTM)

- Pros:
  - Effective for understand the context in sequential data like tweets.
  - Avoiding the vanishing gradient problem
  - Well-suited for sequential data and capturing the temporal dynamics in text
- Cons:
  - Train slower compared to CNNs.
  - With limited data, may overfit

Gated Recurrent Neural Networks (GRU)

Convolutional Neural Networks (CNN)

Bidirectional Encoder Representations from Transformers (BERT)

Discussion

Long Short-Term Memory (LSTM)

Gated Recurrent Neural Networks (GRU)

- Pros:
  - Fewer computational resources.
  - Also good at handling sequential data and capturing temporal dependencies.
- Cons:
  - Limited Capacity, not powerful as LSTM in capturing long-term dependencies.’
  - Similarly to LSTM, potential for overfitting on smaller datasets

Convolutional Neural Networks (CNN)

Bidirectional Encoder Representations from Transformers (BERT)

Presenter: Pan Lu (z5530515)

Discussion

Long Short-Term Memory (LSTM)

Gated Recurrent Neural Networks (GRU)

Convolutional Neural Networks (CNN)

- Pros:
  - Faster train compared to LSTM and GRNN due to the property of parallelism.
  - Good at extracting local features and patterns in text
  - CNN has less risk of overfitting due to shared wights and local connectivity.
- Cons:
  - Lmited Contextual Understanding compared to RNN.
  - Required fixed-size input, limit for variable-length text data.

Bidirectional Encoder Representations from Transformers (BERT)

Presenter: Pan Lu (z5530515)

Discussion

Long Short-Term Memory (LSTM)

Gated Recurrent Neural Networks (GRU)

Convolutional Neural Networks (CNN)

Bidirectional Encoder Representations from Transformers (BERT)

- Pros:
  - Pre-trained representations, leading to improved performance
  - Can capture bidirectional context, better comprehension of the text.
- Cons:
  - Required substantial computational resources, time and memory.
  - Architecture is complex, harder to interpret and fine-tune effectively.

[5000/5000 23:33, Epoch 5/5]

Epoch	Training Loss	Validation Loss	Accuracy
1	0.119700	0.221993	0.934500
2	0.108100	0.267866	0.936500
3	0.050000	0.302585	0.938000
4	0.015000	0.362121	0.934000
5	0.011500	0.376789	0.931500

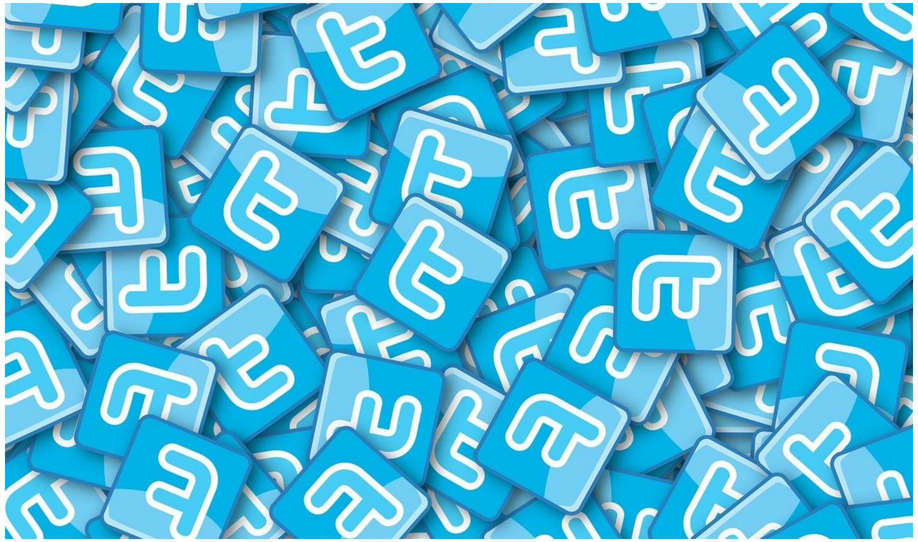
Conclusion

There is no perfect model, only the most suitable model. The choice of model depends on the specific requirements of the project. If performance and accuracy are the primary concerns and computational resources are available, BERT is the best choice. For faster and more efficient training with decent performance, CNNs and GRNNs are suitable. LSTMs offer a balance between performance and capturing long-term dependencies but require more resources than GRNNs and CNNs.

Q&A

UNSW SYDNEY

Thank you



Cong Du (z5504217)

Jiangye Song(z5270075)

Jiayue Tang(z5471262)

Pan Lu (z5530515)

Xiangyu Zheng (z5465193)