# Fake-EmoReact -2021

Team Zulu - 陳韋霖 杜葳葳 張奕廷

# Outline

1. Data preprocessing
2. Method
3. Experiment Result
4. Conclusion

# Data Preprocessing

# Steps

1. Concatenate tweet with each of it's reply as data points
2. Text Processing
   a. Remove non-ascii
   b. Convert emoji into meaningful text, ex: ❤️ → red heart
   c. Clean punctuation and contraction, ex: It’s → It is
   d. Replace URL with "$URL$", ex: https://xxx → $URL$
3. Tokenize, build vocabulary and indices

# Method

# Word Embedding

- GloVe: pre-trained word vectors with 840B tokens, dimension 300
- BERT (cased / uncased): contextual embedding
- Fine-Tuned Embedding in training process

# Models

- CNN
- (Bi-) RNN
- (Bi-) GRU
- (Bi-) LSTM
- BERT, RoBERTa
- Electra

# Tackle Data Imbalance

- Data Pairs in Training Data
  - # Real: # Fake= 31799 : 136722 ≈ 1 : 4
- **Weighted-Loss** during training process

# Ensemble

- **Majority voting**: every individual classifier votes for a class, and the majority wins
- **Soft voting**: sum the predicted **probabilities** for class labels, and predict the class label with the largest sum probability
- **Reply voting**: all data pairs with same source tweet vote for a class, and the majority wins
    - Avg. # reply in training: 4.8 per tweet
    - Avg. # reply in evaluation: **36.9** per tweet

# Experiment Result

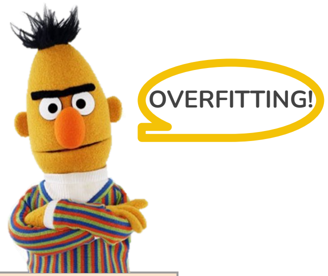
# Training & Practice - Student Track

| Models | BERT(cased) | BERT(uncased) | RoBERTa | Electra |
|---|---|---|---|---|
| F1-score | 0.9777 | 0.9887 | **0.995** | 0.9721 |

| Models | CNN | Bi-RNN | Bi-GRU | Bi-LSTM |
|---|---|---|---|---|
| F1-score | **0.9557** | 0.9006 | **0.9501** | **0.9367** |

| Models | Majority vote | Soft vote |
|---|---|---|
| Bert F1-score | **0.9932** | 0.9929 |
| Non-Bert F1-score | 0.9576 | **0.9579** |

# Training & Practice - Main Track




| Models | BERT(cased) | BERT(uncased) | RoBERTa | Electra |
|---|---|---|---|---|
| F1-score | **0.5495** | **0.5962** | **0.5675** | **0.501** |

| Models | CNN | Bi-RNN | Bi-GRU | Bi-LSTM |
|---|---|---|---|---|
| F1-score | **0.82** | 0.6798 | **0.8099** | **0.7954** |

| Models | Majority vote | Majority vote + Reply vote |
|---|---|---|
| Non-Bert F1-Score | 0.8366 | **0.8703** |

# Evaluation - Student Track

| # | User | Entries | Date of Last Entry | Team Name | Precision score ▲ | Recall score ▲ | F1 score ▲ | Detailed Results |
|---|---|---|---|---|---|---|---|---|
| 1 | ccc_gogo | 8 | 06/02/21 | | 0.8532 (2) | 0.8456 (1) | 0.8435 (1) | View |
| 2 | ChenMian | 8 | 06/01/21 | Team Edward | 0.8399 (3) | 0.8334 (2) | 0.8314 (2) | View |
| 3 | Papa | 11 | 06/01/21 | Team Papa | 0.8300 (4) | [illegible] | [illegible] | [illegible] |
| 4 | Yao | 1 | 06/02/21 | | 0.8560 (1) | 0.8129 (4) | 0.8095 (4) | View |
| 5 | TeamZulu | 15 | 06/02/21 | Team Zulu | 0.8075 (6) | 0.8066 (5) | 0.8060 (5) | View |
| 6 | TeamJuliet | 17 | 06/01/21 | | 0.8002 (7) | 0.7997 (6) | 0.7998 (6) | View |
| 7 | SpencerChen | 5 | 05/31/21 | Team Foxtrot | 0.8215 (5) | 0.7951 (7) | 0.7886 (7) | View |
| 8 | Brett | 13 | 06/01/21 | | 0.7905 (9) | 0.7870 (8) | 0.7855 (8) | View |
| 9 | yuchingtw | 7 | 06/02/21 | Team Victor | 0.7442 (16) | 0.7211 (9) | 0.7162 (9) | View |
| 10 | LuoHeZhou | 5 | 06/01/21 | Team Charlie | 0.7480 (14) | 0.7142 (10) | 0.7016 (10) | View |
| 11 | TeamIndia | 9 | 06/01/21 | Team India | 0.7726 (10) | 0.6981 (11) | 0.6725 (11) | View |
| 12 | Team_Oscar | 4 | 05/30/21 | Team Oscar | 0.7941 (8) | 0.6851 (12) | 0.6490 (12) | View |
| 13 | yiching5417 | 6 | 06/01/21 | Team Mike | 0.6565 (25) | 0.6489 (16) | 0.6432 (13) | View |
| 14 | linzinofan | 8 | 06/02/21 | team November | 0.7394 (18) | 0.6664 (13) | 0.6353 (14) | View |
| 15 | ku4201 | 1 | 05/30/21 | Team Tango | 0.7483 (13) | 0.6647 (14) | 0.6302 (15) | View |

Bi-LSTM Reply Voting

# Evaluation - Main Track

| # | User | Entries | Date of Last Entry | Team Name | Precision score ▲ | Recall score ▲ | F1 score ▲ | Detailed Results |
|---|---|---|---|---|---|---|---|---|
| 1 | Yao | 4 | 06/01/21 | | 0.9346 (1) | 0.9474 (1) | 0.9390 (1) | View |
| 2 | Jina | 10 | 06/02/21 | dx_SKKU+Raon | 0.8427 (2) | [illegible] | [illegible] | [illegible] |
| 3 | SpencerChen | 1 | 05/30/21 | Team Foxtrot | 0.8327 (3) | 0.7937 (4) | 0.8040 (3) | View |
| 4 | TeamZulu | 5 | 06/02/21 | Team Zulu | 0.7993 (5) | 0.7971 (3) | 0.7981 (4) | View |
| 5 | skblaz | 7 | 05/31/21 | | 0.8062 (4) | 0.7077 (5) | 0.7140 (5) | View |

Bi-LSTM

# Conclusion

- CNN and BERT have unexpected performance.
- Ensemble
- Properties of different dataset (train, dev, eval)
  - # of replies per tweet
  - real / fake ratio

# Appendix: Explainable AI - Saliency Map

Saliency score indicates how sensitive the model's final prediction is to each word embedding, which could give us a hint on how much each word embedding contributes to the final decision.

$$w(e) = \frac{\partial(S_c)}{\partial e} \Big|_e \qquad S(e) = |w(e)|$$

We implement the saliency map on our trained **Bi-LSTM** and **CNN** model.

# Appendix: Explainable AI - Saliency Map

bi-lstm

```
predict:      fake
ground truth: fake
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | do | not | believe | that | $URL$ | @criscarter80 | I | aint | No | Fool | unamused | face | unamused | face | fire | fire | fire | #LakeShow | #LAbron | #KobeDaGOAT | $URL$ |
| 1 | 3.545 | 3.793 | 4.919 | 6.594 | 5.275 | 8.139 | 7.271 | 7.108 | 7.343 | 5.915 | 5.357 | 5.455 | 4.899 | 4.964 | 4.301 | 3.489 | 2.826 | 2.515 | 2.216 | 1.947 | 1.353 |

cnn

```
predict:      fake
ground truth: fake
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | do | not | believe | that | $URL$ | @criscarter80 | I | aint | No | Fool | unamused | face | unamused | face | fire | fire | fire | #LakeShow | #LAbron | #KobeDaGOAT | $URL$ |
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 25.397 | 28.415 | 60.958 | 18.666 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

# Appendix: Explainable AI - Saliency Map

bi-lstm

```
predict:      real
ground truth: real
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Today | is | hard | . | My | chest | feels | heavy | . | My | anxiety | is | through | the | roof | . | Verge | of | tears | . | I | want | this | to | end | already |
| 1 | 5.082 | 6.133 | 4.791 | 5.349 | 4.235 | 5.055 | 7.048 | 7.117 | 5.356 | 3.756 | 3.639 | 4.405 | 4.011 | 4.469 | 4.105 | 5.234 | 4.643 | 3.063 | 2.047 | 1.932 | 1.287 | 1.101 | 0.917 | 0.720 | 0.558 | 0.467 |

cnn

```
predict:      real
ground truth: real
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Today | is | hard | . | My | chest | feels | heavy | . | My | anxiety | is | through | the | roof | . | Verge | of | tears | . | I | want | this | to | end | already |
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 18.666 | 13.484 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 28.285 | 21.018 | 7.485 | 5.346 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

# Q & A