AI-Driven Detection of Stress in Social Media Communications

Diamon Dunlap, Yuting Weng, Chen Hui Wang
7.5 minutes/30 slides for 3-person teams

Motivation

кєℓℓαℓєηα
@topaz_kell
Following

Suggested serving size? You don't know me. You don't know what I've been through.

7:48 AM - 23 Nov 2017

242 Retweets 500 Likes

9 242 500

Kyle Weiser
@weiser_thanmost

HAPPY BIRTHDAY EARTH I CANT BELIEVE YOU ARE 2020 YEARS OLD TODAY 😍😍😍 you big beautiful rock

8 4:03 PM - Jan 1, 2020

See Kyle Weiser's other Tweets

Dataset1, Model Training and Validation

Balance Dataset

(8900 in total, 4354 stress, 4366 non stress)

Near-even split between stress-positive and stress-negative tweets

Contextual Insights

Word clouds reveal key terms for each category, included #mentalhealth and #stress for stress group, #happy and #excited for non-stree group

	text	hashtags	labels
0	Being s mom is cleaning 24/7 the same shit ove...	['momlife', 'kids', 'tired']	1
1	And now we have been given the walkthru book b...	['walkthru']	0
2	Wishing YOU Peace Joy & Love! JoyTrain MentalH...	['Peace', 'Joy', 'Love', 'JoyTrain', 'MentalHe...	0
3	speak-no-evil monkey Can I Be Honest With You...	['therapy', 'help', 'NLP', 'CBT', 'hypnotherap...	1
4	Psy Do u hv any regrets? Me No Psy Are you hap...	[]	0

Word cloud for stress group

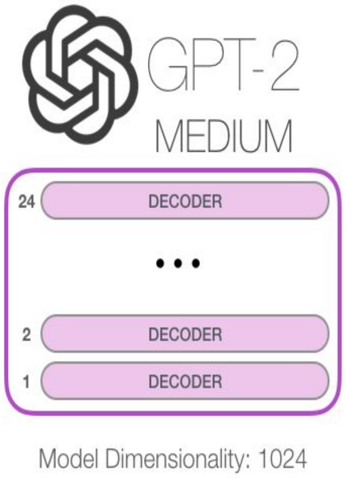
Dataset2, Unlabeled Data for Out of Sample Prediction

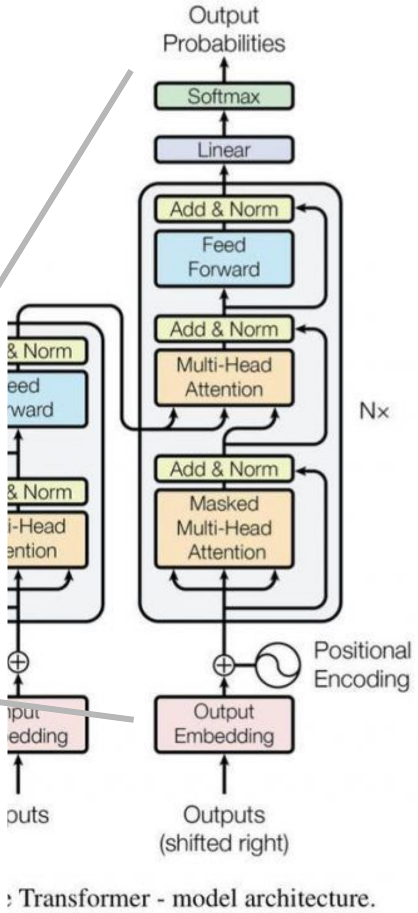
The dataset features several columns that are instrumental for analyzing the social dynamics, reach and engagement for stress and non-stress tweets

post_id	post_created	post_text	user_id	followers	friends	favourites	statuses	retweets
637894677824413696	Sun Aug 30 07:48:37 +0000 2015	It's just over 2 years since I was diagnosed w...	1013187241	84	211	251	837	0
637890384576778240	Sun Aug 30 07:31:33 +0000 2015	It's Sunday, I need a break, so I'm planning t...	1013187241	84	211	251	837	1
637749345908051968	Sat Aug 29 22:11:07 +0000 2015	Awake but tired. I need to sleep but my brain ...	1013187241	84	211	251	837	0
637696421077123073	Sat Aug 29 18:40:49 +0000 2015	RT @SewHQ: #Retro bears make perfect gifts and...	1013187241	84	211	251	837	2
637696327485366272	Sat Aug 29 18:40:26 +0000 2015	It's hard to say whether packing lists are mak...	1013187241	84	211	251	837	1

Model 1 - Fine-tuned GPT-2

We choose to use a GPT model despite it being a unidirectional, generative model to see if we could leverage its transfer learning.

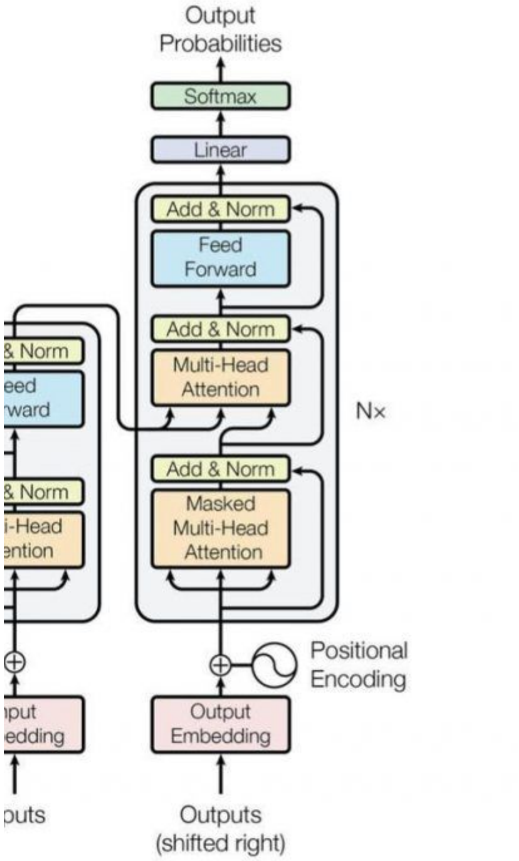




Model 1 - Fine-tuned GPT-2

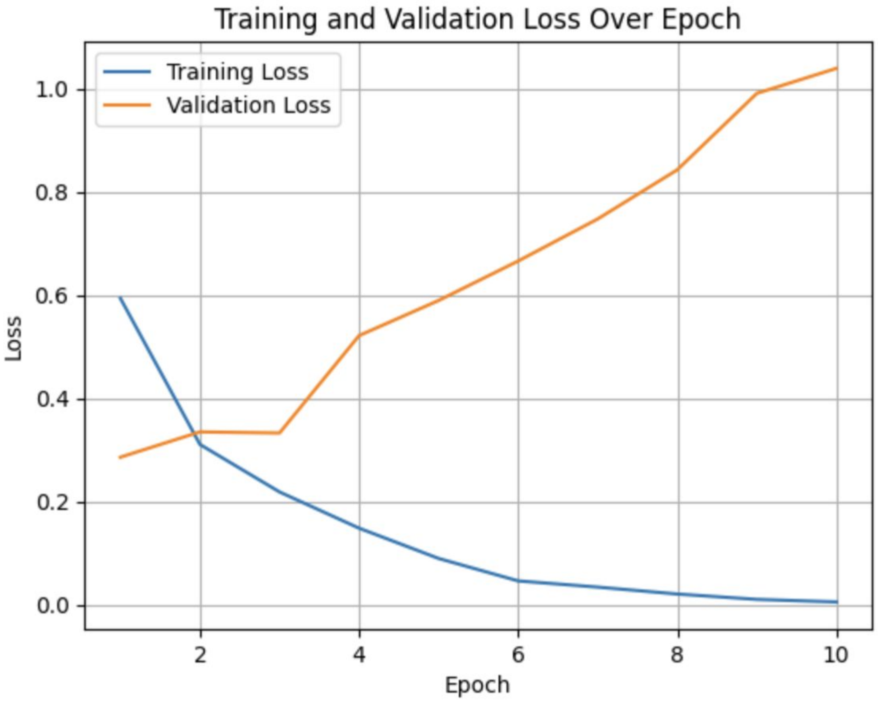
We choose to use a GPT model despite it being a unidirectional, generative model to see if we could leverage its transfer learning.

Our hypothesis: BERT-based models designed for next sequence prediction and text classification will outperform this GPT-2 model.



: Transformer - model architecture.

Model Parameters and Training



Model parameters - batch size 8, weight decay 0.01, adam optimizer

Validation and Testing Results

Validation Metrics

After 3 epochs:

Accuracy : 0.8708

Precision : 0.9071

Recall : 0.8281

F1 : 0.8706

Testing Metrics

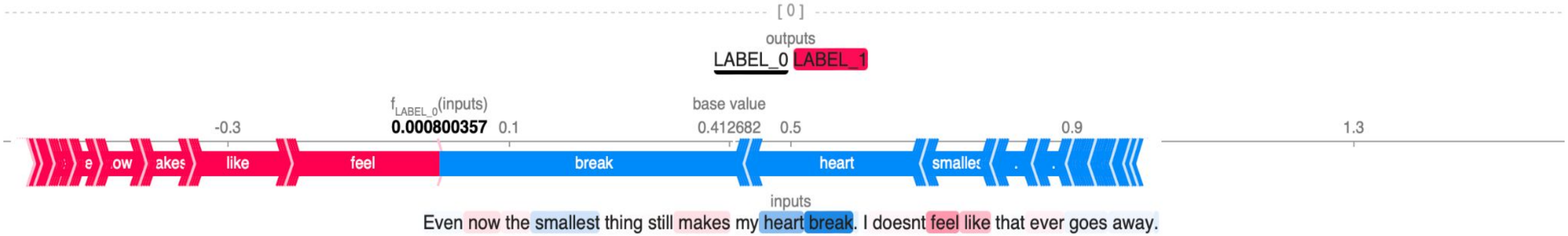
Accuracy : 0.8607

Precision : 0.8571

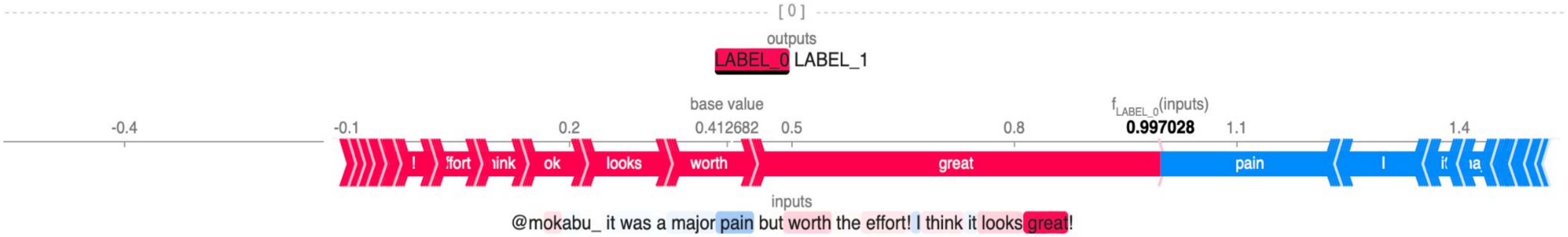
Recall : 0.8686

F1 : 0.8607

Feature Importance with Shap Values

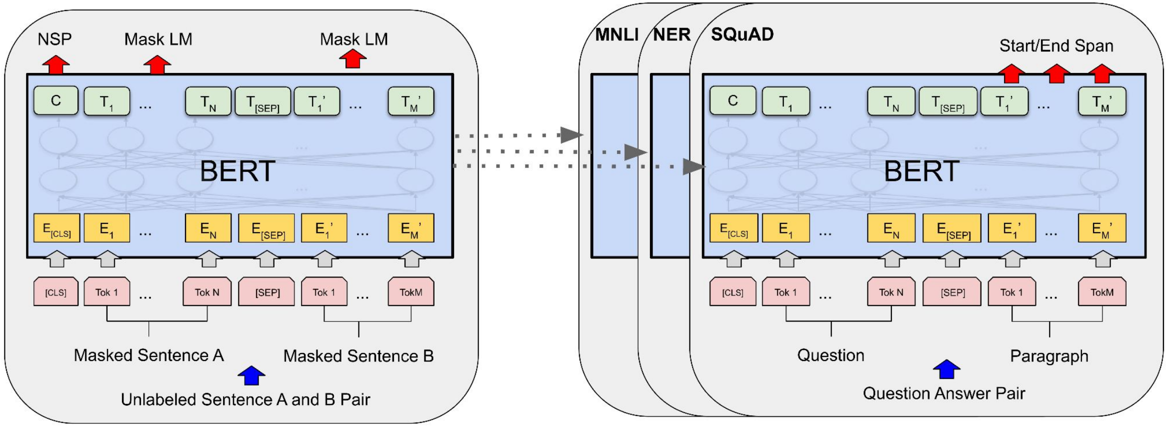



```
ts.text(shap_0)
```



Model 2 Pre-trained BERT

Unlike GPT-2's sequential generation capabilities, we assume that BERT is more suitable than GPT2 in the task because its designed for text classification tasks.



Pre-training

Fine-Tuning

Data Splitting

Training/Validation

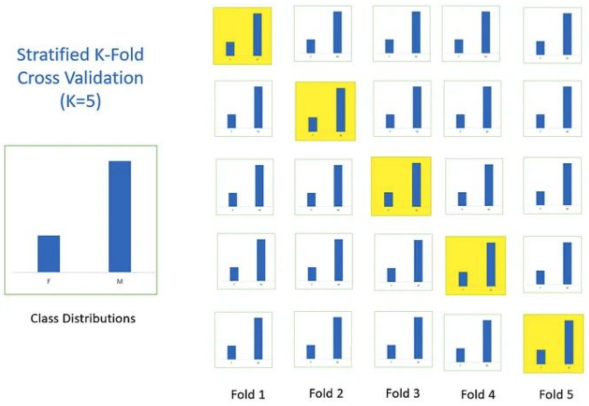
90% of the dataset was used for training and validation, ensuring a balanced distribution of stress and non-stress labels.

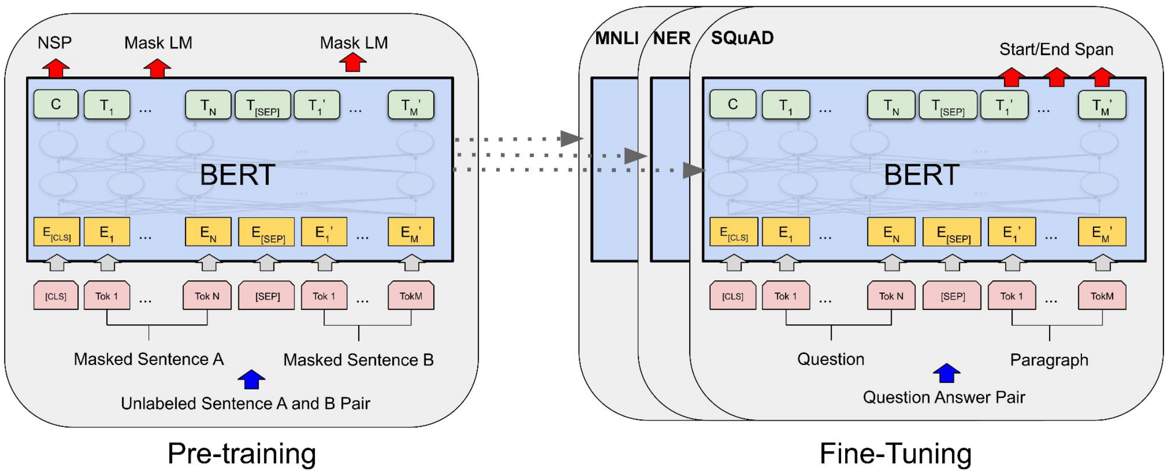
Testing

The remaining 10% was reserved for independent testing to evaluate the model's performance.

StratifiedKFold + Pre-trained BERT

First employs StratifiedKFold for data splitting across training and validation sets; then trains the BERT model on each split of the data, assessing its performance on the validation set at the end of each training epoch





Model Evaluation

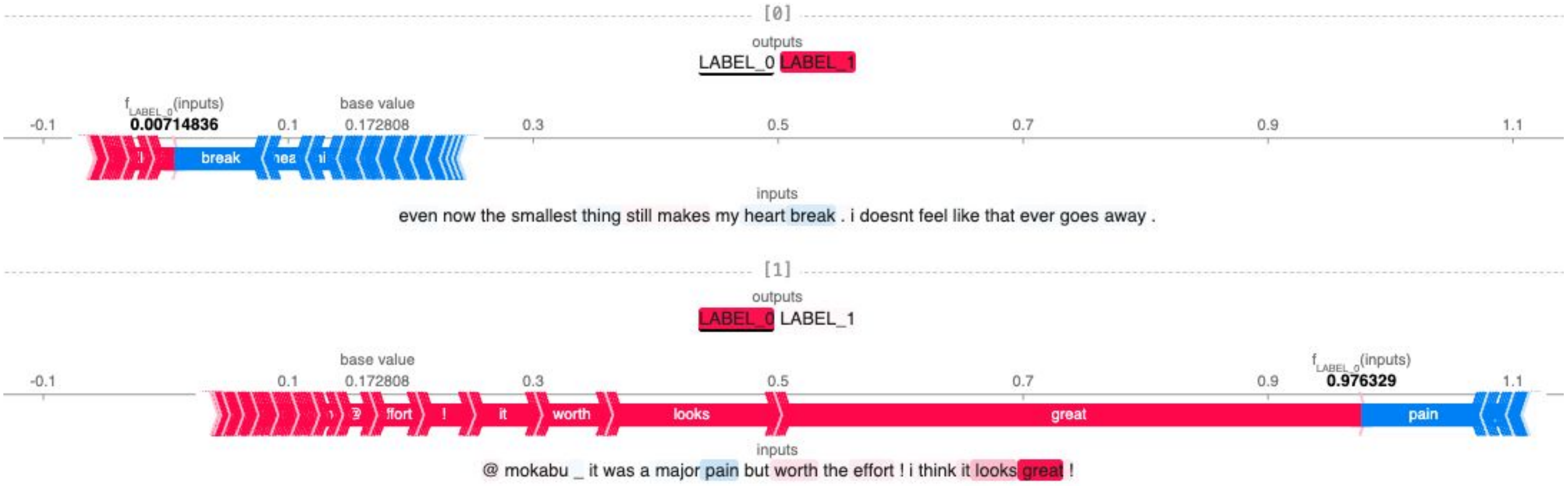
Validation Metrics

The model consistently achieved over 85% accuracy, precision, recall, and F1 score.

Testing Metrics

Accuracy : 0.8977
Precision : 0.8986
Recall : 0.9006
F1 : 0.8996

Shapley Value



Model 3: LSTM or CNN

LSTM (Long Short-Term Memory) and CNN (Convolutional Neural Network) models are traditional deep learning approaches for text classification.

LSTM: capture long-term dependencies in sequential data.

CNN: capture local patterns such as key phrases or n-grams.

We hypothesize that the **LSTM model will perform better** due to its ability to capture contextual information in sequences of text.

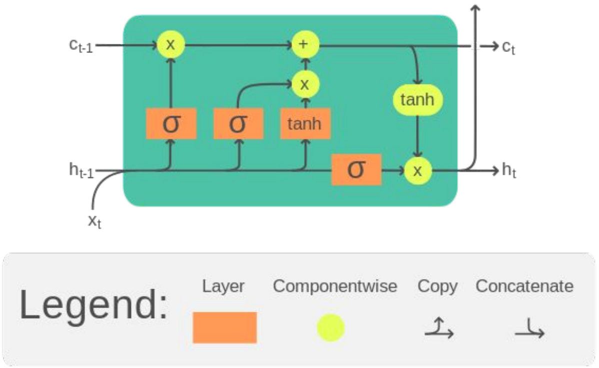
Building and Training

Step 1: Model Architecture

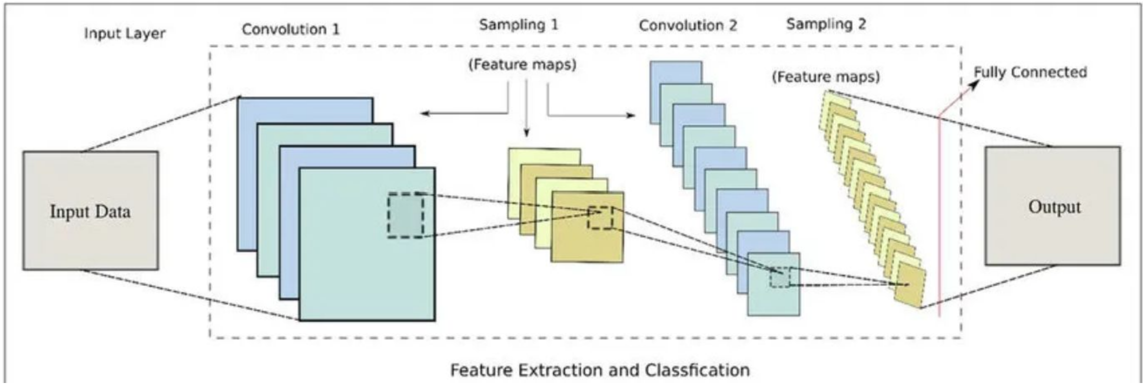
Step 2: Data Splitting

Step 3: Model Training: Stratified K-Fold cross-validation with 5 splits, and EarlyStopping

LSTM



CNN



Model Evaluation

Validation Metrics

LSTM
Loss: 0.968
Accuracy: 0.804
Precision: 0.788
F1 Score: 0.814

CNN
Loss: 0.428
Accuracy: 0.832
Precision: 0.841
F1 Score: 0.833

Testing Metrics

LSTM
Loss: 1.126
Accuracy: 0.782
Precision: 0.770
F1 Score: 0.792

CNN
Loss: 0.501
Accuracy: 0.806
Precision: 0.811
F1 Score: 0.809

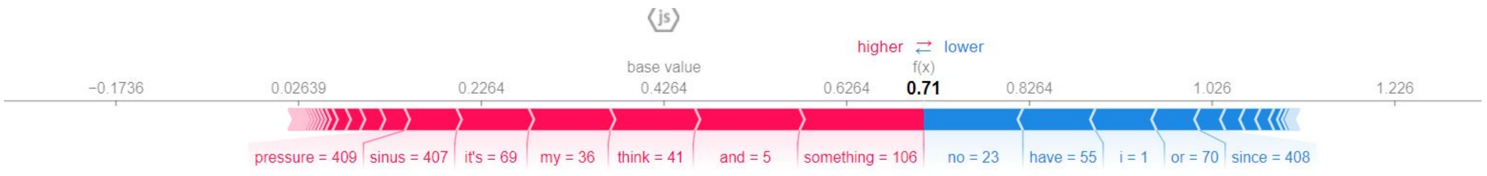
Comparison between LSTM and CNN

The CNN model outperforms the LSTM model across all evaluated metrics.

Possible Reasons:

- **Text Length**
- **Local Patterns**
- **Dataset Size**

Shapley Value Example: “@MissLusyd I have no idea. My throat hurts too. I think it's dry sinus or something since I feel pressure and junk.”



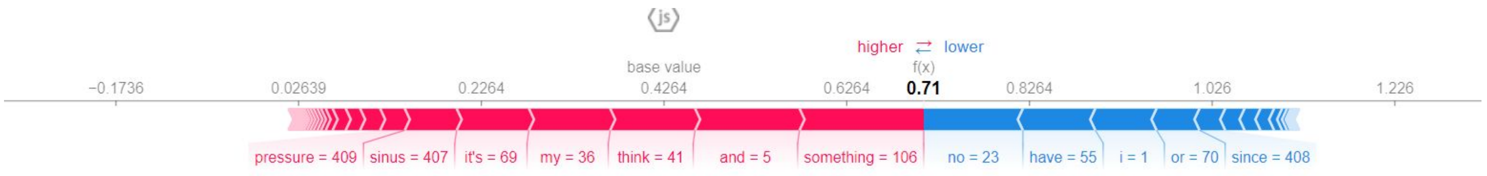
Comparison between LSTM and CNN

The CNN model outperforms the LSTM model across all evaluated metrics.

Possible Reasons:

- **Text Length**
- **Local Patterns**
- **Dataset Size**

Shapley Value Example: “@MissLusyd I have no idea. My throat hurts too. I think it's dry sinus or something since I feel pressure and junk.”



Conclusion - Feature importance amongst all models

Label 0: shap values for all 3 models top 10 separating by label

BERT

	Word	Label	SHAP Value
545	positive	LABEL_0	0.787985
96	appreciate	LABEL_0	0.564489
129	best	LABEL_0	0.453644
753	welcome	LABEL_0	0.430993
287	great	LABEL_0	0.422258
675	thank	LABEL_0	0.405098
296	happy	LABEL_0	0.389967
181	cool	LABEL_0	0.376580
718	true	LABEL_0	0.366154
696	thrill	LABEL_0	0.328247

GPT-2

	Word	Label	SHAP Value
535	positive	LABEL_0	0.906217
849	Important	LABEL_0	0.787853
91	FUN	LABEL_0	0.776189
614	thank	LABEL_0	0.765982
289	bright	LABEL_0	0.756163
264	appreciate	LABEL_0	0.715291
100	Great	LABEL_0	0.696920
105	Happy	LABEL_0	0.659856
354	favorite	LABEL_0	0.635546
55	Best	LABEL_0	0.634769

CNN

	Word	SHAP Value	Label
186	fukc	-0.002934	0.0
403	positive	-0.002128	0.0
421	recognize	-0.001839	0.0
367	ones	-0.001815	0.0
227	history	-0.001493	0.0
168	favourite	-0.001448	0.0
26	already	-0.001112	0.0
382	part	-0.001051	0.0
221	henhouse	-0.000974	0.0
570	vaccines	-0.000970	0.0

Conclusion - Feature importance amongst all models

Label 1: out of sample analysis for all models

BERT

	Word	Label	SHAP Value
645	strange	LABEL_1	0.612492
467	murder	LABEL_1	0.453731
219	drunk	LABEL_1	0.405905
350	ins	LABEL_1	0.392226
885	ecure	LABEL_1	0.373147
60	abusive	LABEL_1	0.367399
557	quote	LABEL_1	0.336703
333	hungry	LABEL_1	0.330707
413	literally	LABEL_1	0.310816
683	therapist	LABEL_1	0.301211

GPT-2

	Word	Label	SHAP Value
598	strange	LABEL_1	0.790881
419	insecure	LABEL_1	0.718702
324	die	LABEL_1	0.622088
533	pointless	LABEL_1	0.530606
257	animals	LABEL_1	0.491791
335	drunk	LABEL_1	0.474953
462	lost	LABEL_1	0.440963
325	different	LABEL_1	0.375023
1024	blems	LABEL_1	0.331876
646	tri	LABEL_1	0.327656

CNN

	Word	SHAP Value	Label
499	sunglasses	0.004673	1.0
522	their	0.004610	1.0
92	cartoon	0.004210	1.0
66	become	0.004047	1.0
191	gay	0.004009	1.0
95	cause	0.004007	1.0
343	naturopathy	0.003912	1.0
391	period	0.003850	1.0
498	sunday	0.003605	1.0
496	summoned	0.003195	1.0

Conclusion - Twitter Engagment Behavior (BERT)

Followers & Friends

Tweets predicted as non-stress have more followers and Friends in average

Favorites

Non-Stress tweets also have higher mean favorites

Retweets

Both groups have no retweets count by median

Sentiment Score

non-stress group has a higher sentiment score (more positive)

predicted_label	
LABEL_0	24
LABEL_1	76

predicted_label	followers	friends	favourites	retweets	textblob_sentiment
LABEL_0	1865.333333	1526.958333	8968.791667	0.0	0.253310
LABEL_1	846.026316	692.486842	7534.210526	0.0	0.025887

Conclusion - Twitter Engagment Behavior (BERT)

Followers & Friends

Tweets predicted as non-stress have more followers and Friends in average

Favorites

Non-Stress tweets also have higher mean favorites

Retweets

Both groups have no retweets count by median

Sentiment Score

non-stress group has a higher sentiment score (more positive)

predicted_label	
LABEL_0	24
LABEL_1	76

predicted_label	followers	friends	favourites	retweets	textblob_sentiment
LABEL_0	1865.333333	1526.958333	8968.791667	0.0	0.253310
LABEL_1	846.026316	692.486842	7534.210526	0.0	0.025887

Limitation - Biases in Auto-labeled Data

Oversimplified Categorization

Human emotions are complex, but the labeled dataset may reduce them to overly simplistic labels.

Lack of Generalizability

Models trained on biased data may fail to capture context and subtlety in emotional expression and struggle to apply to diverse populations and new situations

Limitation - Biases in Auto-labeled Data

Oversimplified Categorization

Human emotions are complex, but the labeled dataset may reduce them to overly simplistic labels.

Lack of Generalizability

Models trained on biased data may fail to capture context and subtlety in emotional expression and struggle to apply to diverse populations and new situations

Limitation - Differences in Stress Expression

Alternative modes of emotional expression, such as through **images** and symbols.

Stress across **diverse cultures** is crucial for effective mental health support.

Future Work

Data

- Manually label out-of-sample test set and see if labeling aligns
- Multimodal modeling by including images and/or video (Instagram)

Modeling

- Discern temporary versus persistent states of stress
- Handle multiple mental disorders and their intersections
- Look at different social media platforms and explore the differences between users
- Ensemble multiple models that handle different mental disorders
- Graph user networks

Application

- Research ethically implemented applications

Future Work

Data

- Manually label out-of-sample test set and see if labeling aligns
- Multimodal modeling by including images and/or video (Instagram)

Modeling

- Discern temporary versus persistent states of stress
- Handle multiple mental disorders and their intersections
- Look at different social media platforms and explore the differences between users
- Ensemble multiple models that handle different mental disorders
- Graph user networks

Application

- Research **ethically implemented applications**

Thank you!