

Ling 573 A: Project

Predicting Human Emotion: Empathy and Distress

by – Divy Sharma, Alonso Ma, Nora Goldfine, and Manisha Singh



Agenda

- Motivation
- Overall Project Architecture
- Dataset overview
- Architecture overview
- Initial system D#2
 - Model and Training
 - Results and Error analysis
- Revised system D#3
 - Lexicon
 - Hyperparameter -tuning
 - Data sampling
 - Result and Error analysis
- References

Motivation

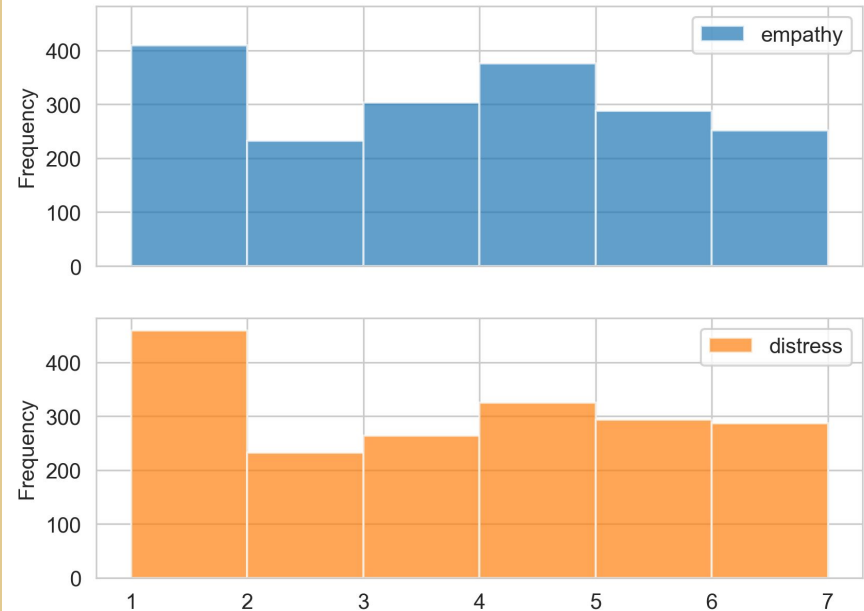
Primary Task: WASSA 2022 Shared Task on Empathy Detection and Emotion Classification: First Subtask - Predict Empathy Concern and Personal Distress at an essay-level

Adaptation Task: WASSA 2023 Shared Task on Empathy Emotion and Personality Detection in Interactions: Track 1 - Empathy and Emotion Prediction in Conversations

Dataset

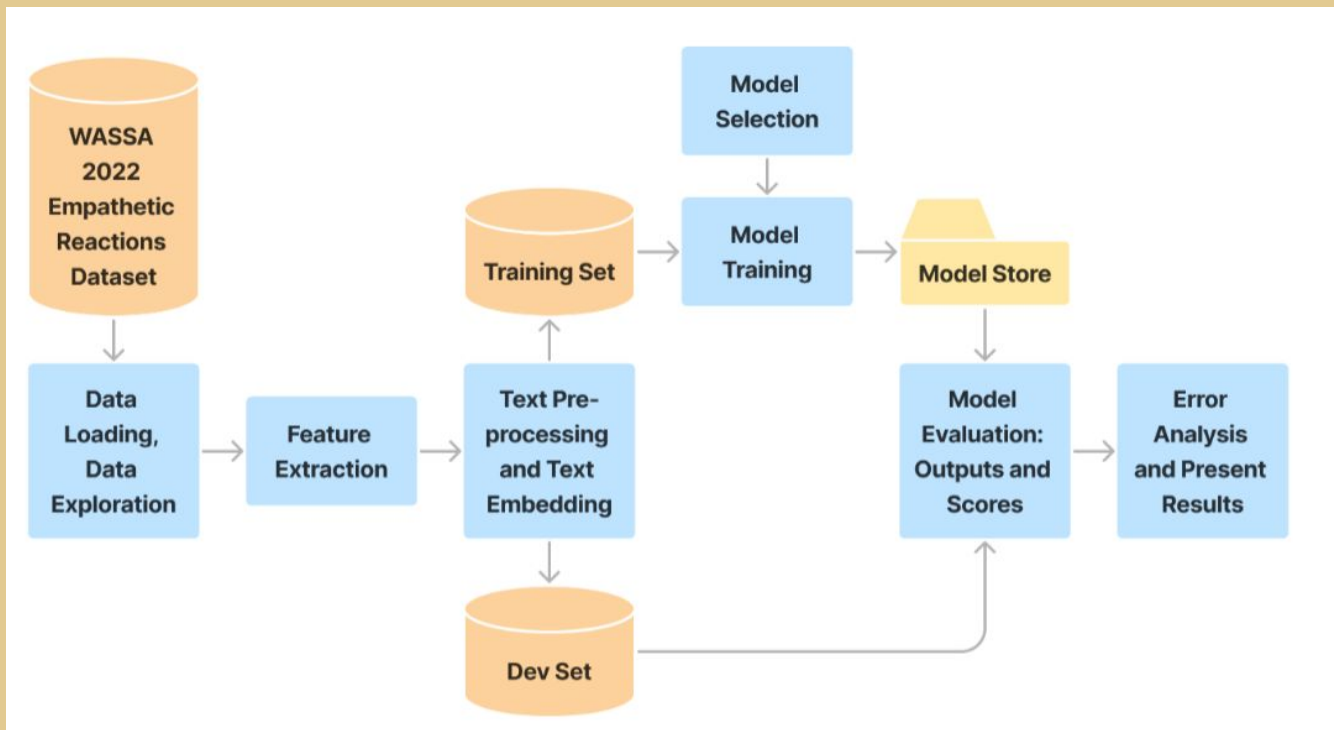
- Training dataset is 1860 rows
- Dev set 270 rows
- Three columns empathy, distress and the essay text

Distribution of Empathy and Distress values in the Training dataset



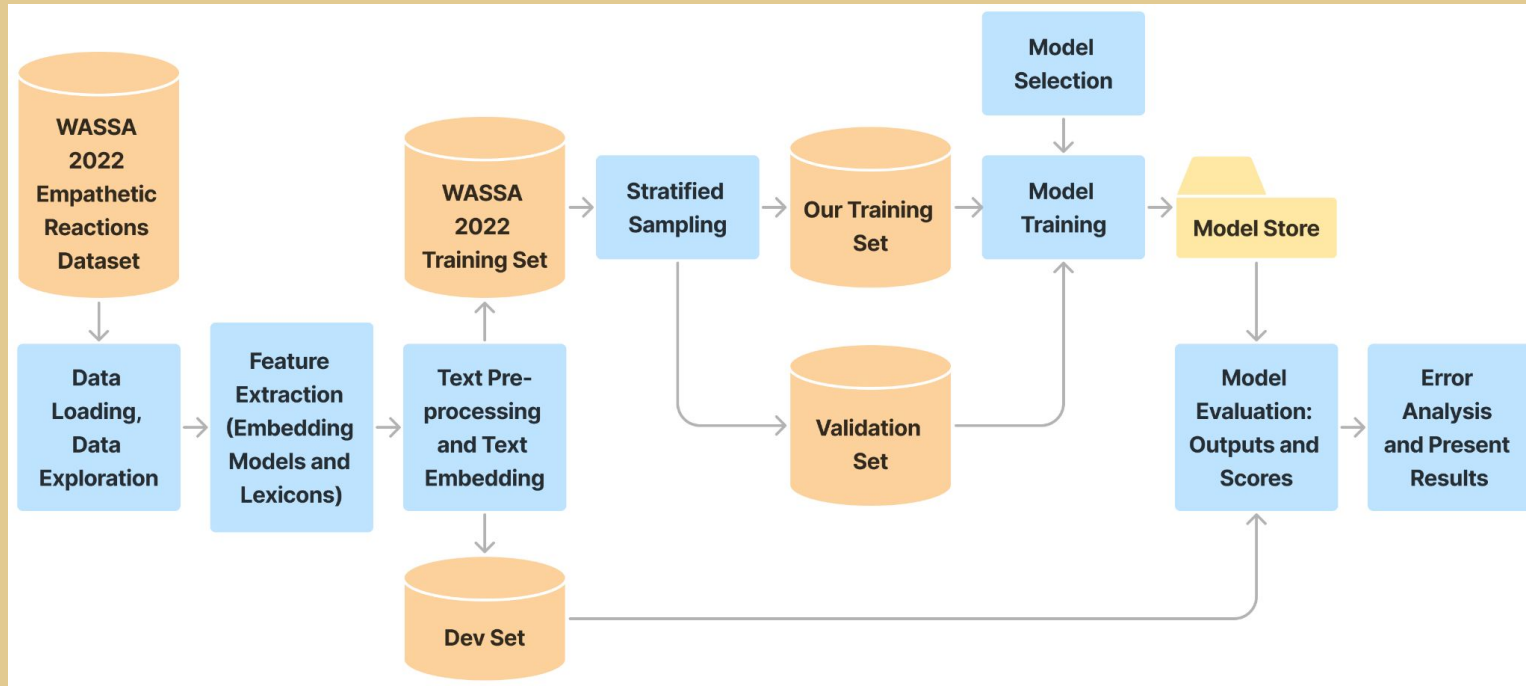
Overall Project Architecture

D#2:



Overall Project Architecture

D#3 with three revisions:



D#2: Initial system

-
- #1: Sentence embedding
 - #2: Model and training
 - #3: Result and error analysis



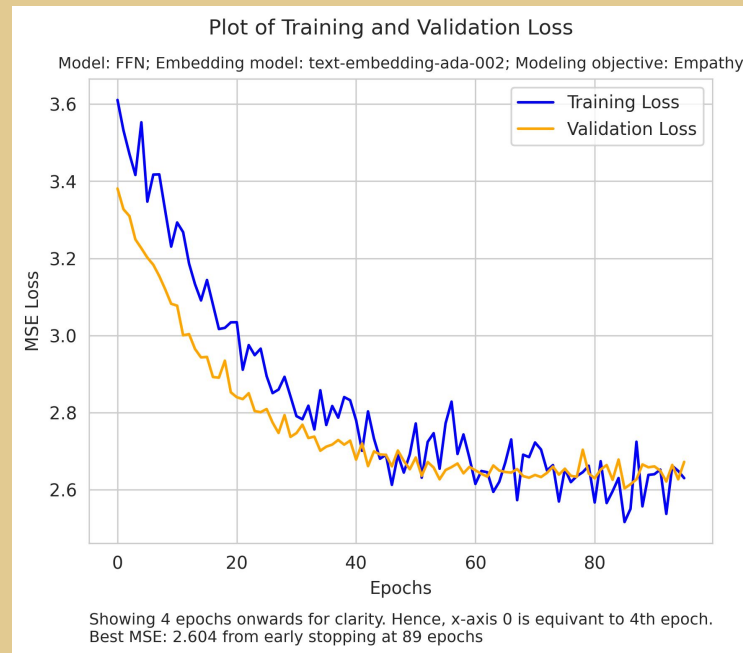
Text processing and Text embedding

	empathy	distress	essay	essay_emb
0	5.667	4.375	it is really diheartening to read about these ...	[0.04520609974861145, 0.058606069535017014, -0...
1	4.833	4.875	the phone lines from the suicide prevention li...	[-0.0877808928489685, 0.009987352415919304, 0....
2	5.333	3.500	no matter what your heritage, you should be ab...	[0.01799248717725277, 0.04691638424992561, -0....
3	4.167	5.250	it is frightening to learn about all these sha...	[0.056801456958055496, -0.0004722552839666605,...
4	5.333	4.625	the eldest generation of russians aren't being...	[0.013306875713169575, 0.04898981750011444, 0....

Embedding Model	Dimensions	Tokenization and Tokens
"all-MiniLM-L6-v2" Sentence Transformer	384 dimensional dense vector	Word Piece tokenization 128 tokens
"all-mpnet-base-v2" Sentence Transformer	768 dimensional dense vector	Word Piece tokenization with 384 tokens
"all-roberta-large-v1" Sentence Transformer	1024 dimensional dense vector	Word Piece tokenization with 128 tokens
OpenAI "text-embedding-ada-002"	1536 dimensional dense vector	BPE tokenization with 8191 tokens

Model and Training

- Sentence level embedding as features
- 20% of the training set data is used as validation set
- FFN using PyTorch
 - two hidden layer (256 and 128 units)
 - ReLU activation
 - Dropout layer $p(0.5)$: to address overfitting
 - MSE loss function
 - Adam as optimizer
 - 100 epochs
- Save the best weights for evaluation



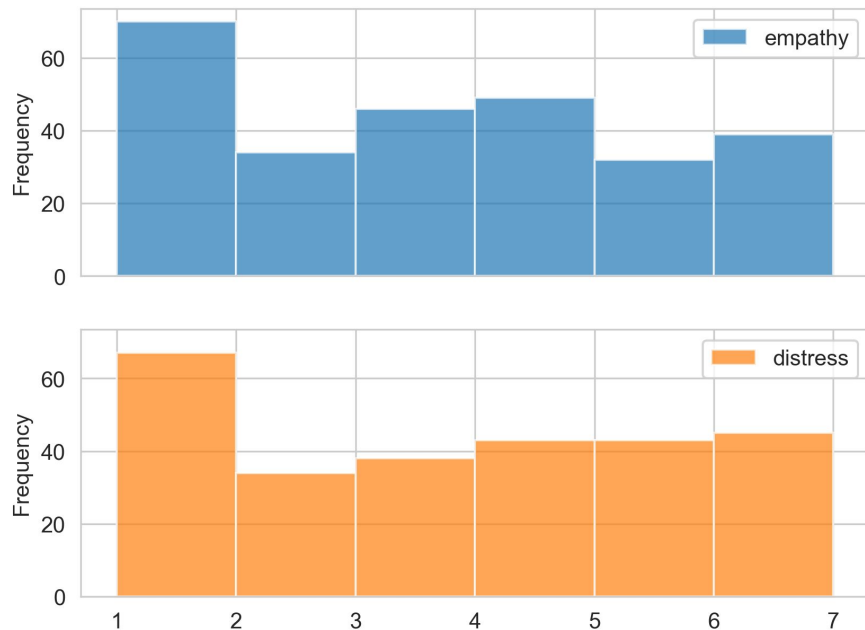
D#2 Results

	Empathy	Distress	Mean
FNN baseline	.379	.401	.390
FNN with all-MiniLM-L6-v2 embedding	.379	.370	.375
FNN with all-mpnet-base-v2 embedding	.386	.324	.355
FNN with all-roberta-large-v1 embedding	.395	.360	.378
FNN with text-embedding-ada-002 embedding	.438	.426	.432

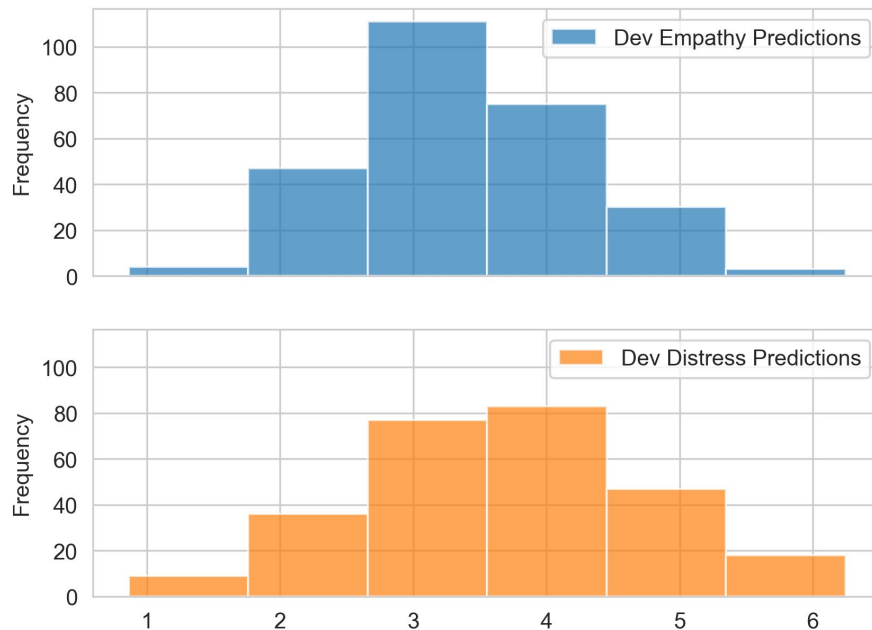
The best result was given by OpenAI “text-embedding-ada-002” , followed by “all-roberta-large-v1” Sentence Transformer”

D#2 Results

Distribution of Empathy and Distress values in the Dev set



Distribution of Empathy and Distress values in the prediction
Model: FFN; Embedding model: text-embedding-ada-002



Distribution of Empathy and Distress values in (a) Dev dataset (b) Predictions

D#2 Error analysis - Manual data check

Example	Gold standard Distress	Predicted Distress #D2	Predicted Distress #D3	Anticipated Reason
I think people would have a field day over this in blogs and forums, especially the conspiracy theorist who thinks some cabal is somewhere controlling events and outcomes, the good news is the pilot survived and was recovered safely and maybe Russia should be more focused on ensuring safety of its military .	5.5	1.41	1.15	Empathy mixed with sarcasm
The way that the United Kingdom is treating its refugees within the camps is nothing short of barbaric. It is actually so sad seeing these people who ran from war get to countries just to be unwanted and hoarded up into camps. The children within the Calais camp in particular are just being neglected and abused and are lacking so many things required for growth	1.87	4.75	5.43	Annotators demography or sarcasm

Example	Gold standard Empathy	Predicted Empathy #D2	Predicted Empathy #D3	Anticipated Reason
When I first started reading this article it made me mad at heroin addicts and I thought it must be these forgetful types nodding off that was causing the crisis to occur. I'm pretty surprised that the name culprit seems to be parents that don't control their medicine cabinets. Somehow, I think that makes me even more angry.	7	2.799	2.71	Empathy with anger
I'm torn on the whole ivory trade situation. I get that farmers and people in their territory are often bothered by elephants (i.e. through crop destruction, etc.), but it doesn't seem fair to cause so much suffering to these animals. I feel bad that the locals are likely poor and feel that they have little choice, but I hate to think of elephants suffering.	1.33	4.65	4.71	Empathy with anger

(Paper reference : [Kuijper et al. 2018](#))

D#3: Project revisions

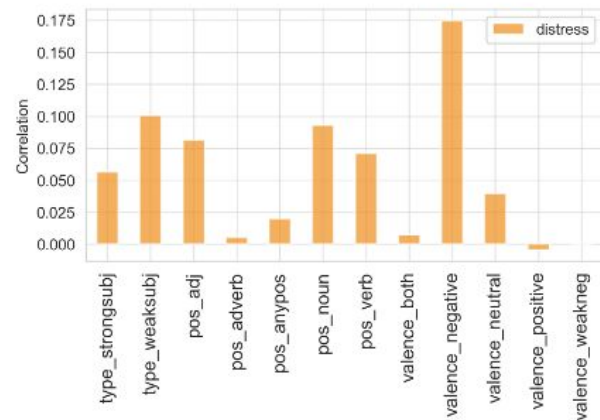
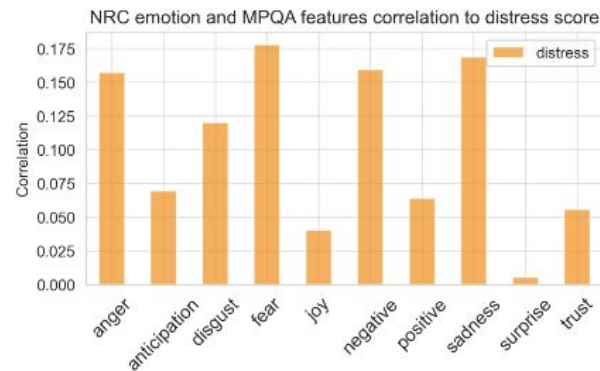
-
- Revision #1: hyperparameter tuning
 - Revision #2: data sampling
 - Revision #3: lexicon features



Lexicon features

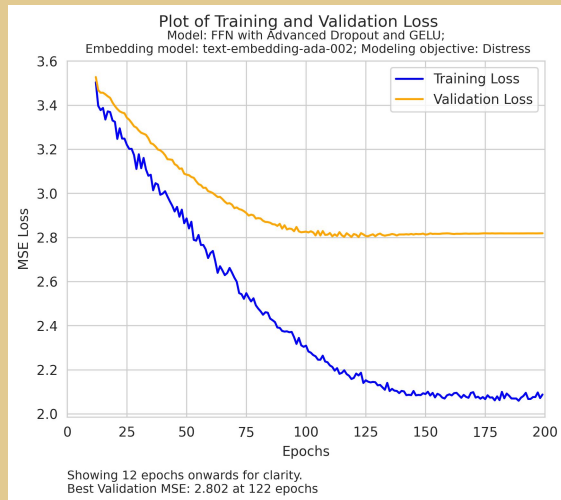
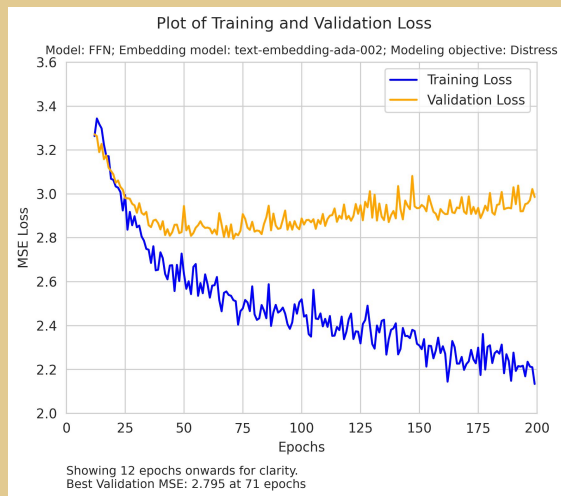
Additional features created based on word counts and mean scores from **4 different lexicons**:

- **NRC emotion association**: 8 basic emotions (Mohammad and Turney, 2013)
- **MPQA subjectivity lexicon**: valence and subjectivity at word level (Wilson et al, 2005)
- **NRC VAD lexicon**: valence, arousal and dominance scores (Mohammad, 2018)
- **Verbal polarity shifters**: list of words that cause shift in positive-negative polarity (Schulder et al, 2018)



Hyperparameter tuning

- **D#2:**
Overfitting in early epochs → dropout included to mitigate
- **D#3:**
 - Advanced dropout (Xie et al, 2021) implemented to **adaptively tune the dropout rate** hyperparameter
 - Pytorch scheduler added to dynamically adjust learning rate when finding a plateau

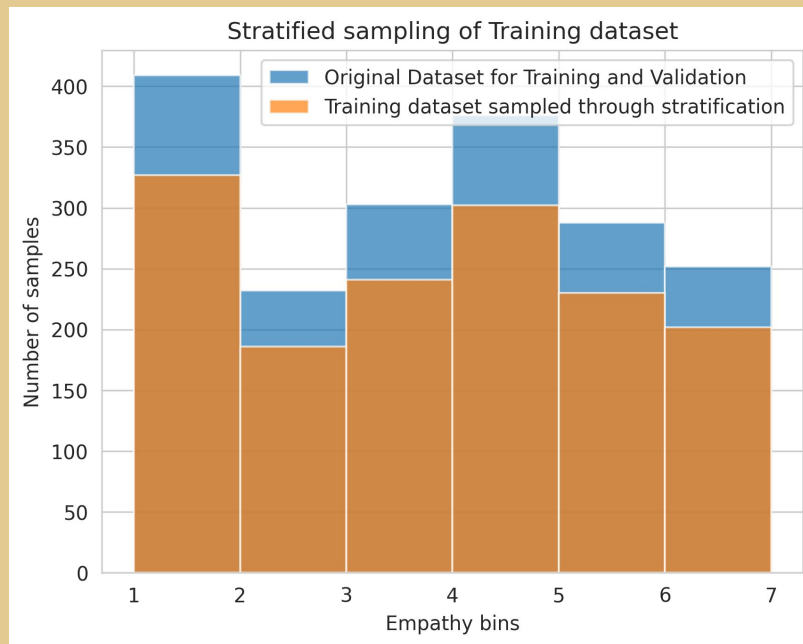


Data sampling

Imbalance in target variable (distribution with higher concentration in lower scores)

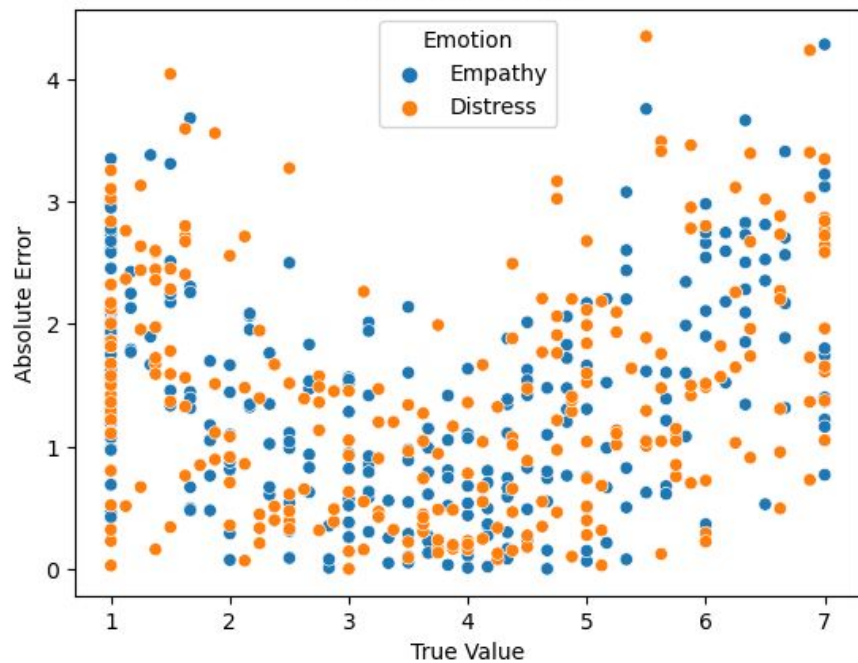
Utilized stratified sampling in Scikit-learn to ensure preservation of target variable distribution

- Since variable is continuous, **bucketization** was carried out beforehand

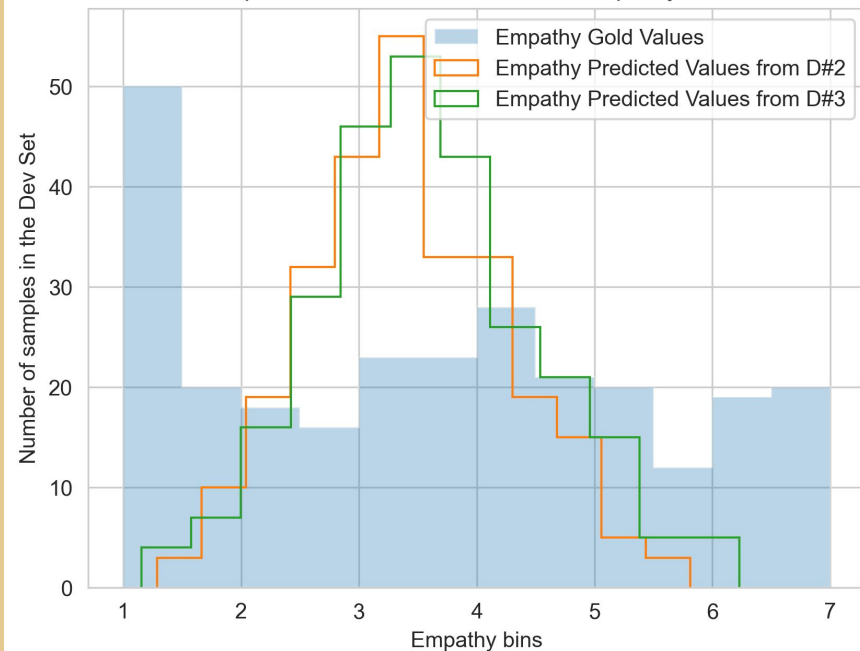


Error analysis

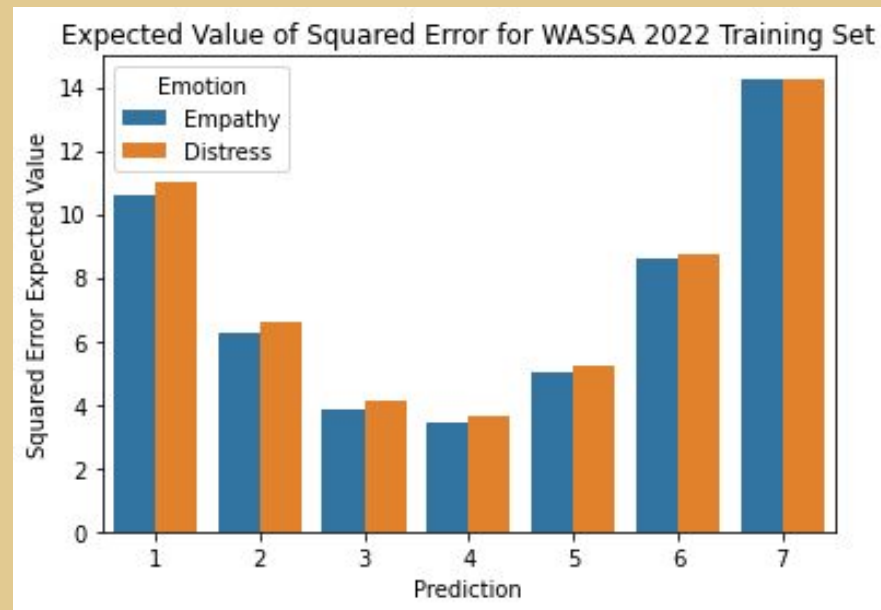
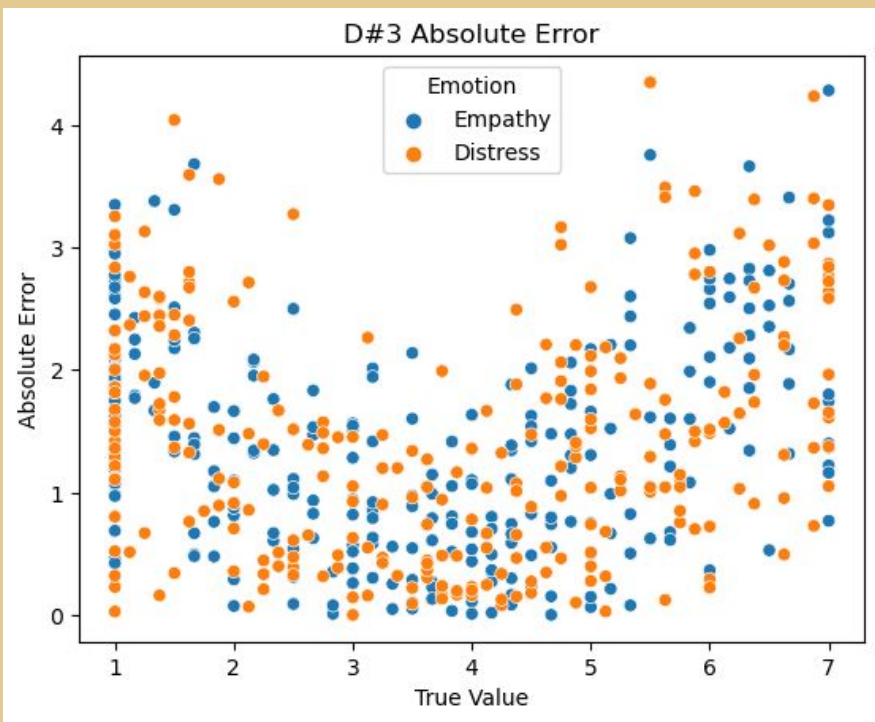
D#3 Absolute Error



Comparison of the Distribution of Empathy Values



Error analysis



Results

	Empathy	Distress	Mean
FNN baseline	.379	.401	.390
FNN best performing model from D#2	.438	.426	.432
D#3 Revision #1: FNN, Advanced Dropout, GELU	.434	.425	.430
D#3 Revision #2: FNN, Advanced Dropout, GELU, and SDS	.475	.448	.462
D#3 Revision #3: FNN, Advanced Dropout, GELU, Lexicons, and SDS	.481	.456	.469

Table 2: Table of results for D#3: Advanced Dropout, GELU, Lexicons, and Stratified Data Sampling (SDS). All results beyond the baseline are with the `text-embedding-ada-002` embedding model

Key References

Motivation, Shared task and Dataset

- [Picard \(2000\)](#)
- [WASSA \(2022\)](#)
- [Buechel et al., 2018](#)
- [WASSA \(2023\)](#)
- [Barriere et al., 2022](#)

Lexicon

- [Mohammad and Turney, 2013](#)
- [Wilson et al., 2005](#)
- [Mohammad, 2018](#)
- [Schulder et al., 2018](#)

Error analysis

- [Kuijper et al. 2018](#)

Embeddings

- [Wang et al., 2020](#)
- [Song et al., 2020](#)
- [Liu et al., 2019](#)
- [Neelakantan et al., 2022](#)
- [Azure-OpenAI, 2023](#)

Hyperparameter tuning

- [Xie et al., 2021](#)
- [Hendrycks and Gimpel \(2020\)](#)



Thank You!



UNIVERSITY *of* WASHINGTON

