[Task Two] - BERT Model (What and Why?)

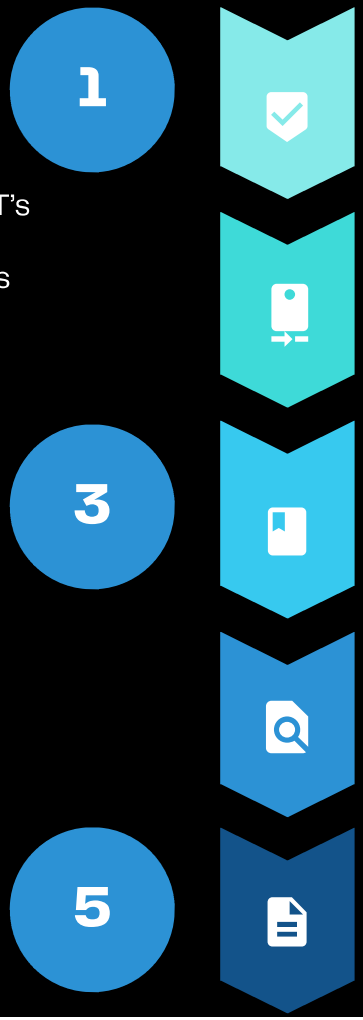
What:

- Bidirectional Encoder Representations from Transformers (BERT) makes use of a transformer, which is an attention mechanism that learns contextual relations between words in a text
- It is a pre-trained deep-learning language model that helps computers understand text using surrounding text to establish context

Why:

- BERT shines when you feed it with large amounts of data, which we have (~100k text reviews)
- Can be more effective than other traditional sentiment analysis models especially when it comes to learning contextual and positional relationships between words like in text reviews
- BERT allows us to handle these subtle contextual cues that plague text reviews

[Task Two] - BERT Model Process Overview



1

Loading & Pre-processing

- Read data into the data frame
- Clean the dataset by removing punctuations but don't stem or remove stopwords
- Pre-process and tokenize using BERT's own tokenizer
- Split into training, validation, and test sets

2

Preparatory Phase

- Import key packages like "Transformers" and "Pytorch"
- Prepare the training and validation data for model creation by encoding them

3

Model Creation & Training

- Create the base model and prepare for model training
- Train the model and wait

4

Choosing the Best Model

- Choose the best model hyperparameter (epoch) based on validation metrics
- Analyse the validation errors and restart if necessary

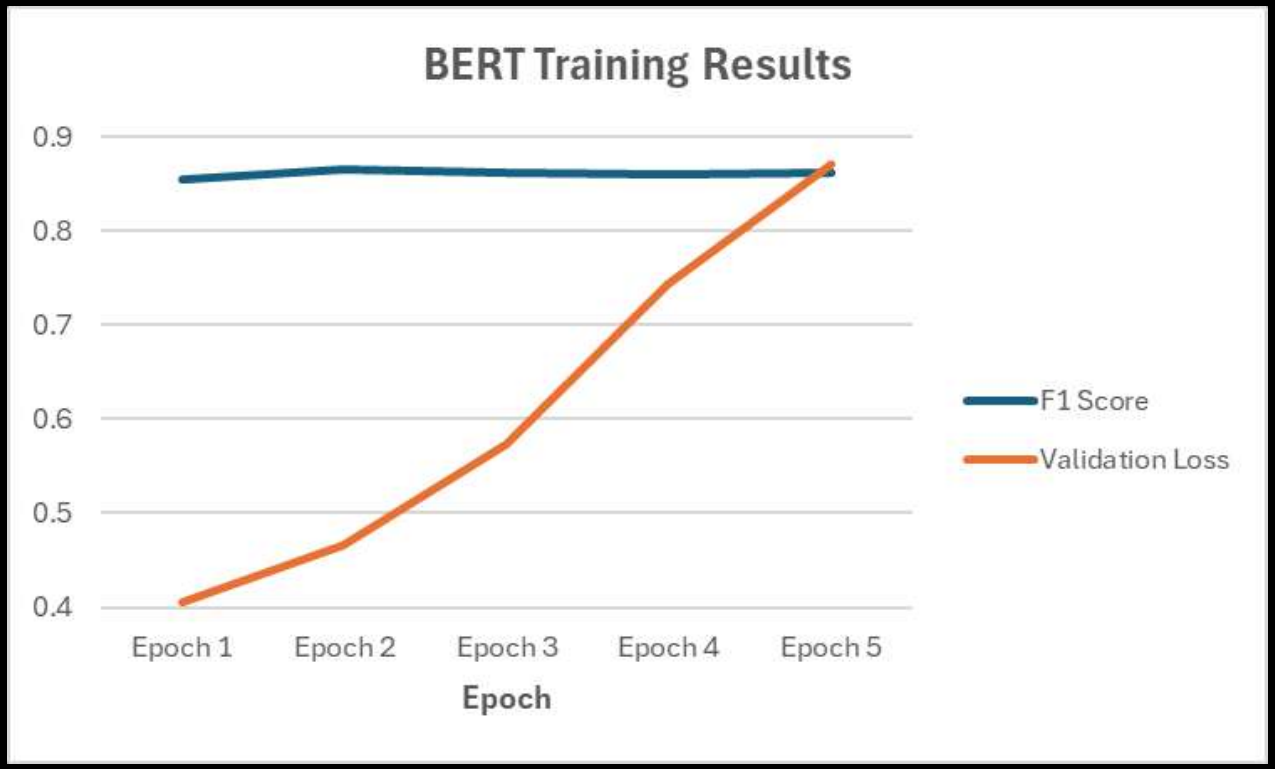
5

Testing & Evaluating the Model

- Test the model using the unseen test dataset
- Evaluate the model by creating a classification report and confusion matrix
- Finalise and start using model

[Task Two] - BERT Model (Choosing Hyperparameter)

Choosing the Optimal Epoch (BERT Hyperparameter)



How to choose? Maximize F1 Score and Minimize Validation Loss.

An epoch in machine learning means one complete pass of the training dataset through the algorithm.

```
Epoch 1
Training loss: 0.4230179212099308
Validation loss: 0.4052461648963951
F1 Score (Weighted): 0.8542600851401484

Epoch 2
Training loss: 0.34474913491362597
Validation loss: 0.46566380172418703
F1 Score (Weighted): 0.865951892072186

Epoch 3
Training loss: 0.26364536303943653
Validation loss: 0.5726030811991109
F1 Score (Weighted): 0.8617343162069128

Epoch 4
Training loss: 0.16909867958677674
Validation loss: 0.7437222844298226
F1 Score (Weighted): 0.860736309521918

Epoch 5
Training loss: 0.09076723472180567
Validation loss: 0.870957531200748
F1 Score (Weighted): 0.8625517707711815
```

Epoch = 2

[Task Two] - BERT Model (Choosing Hyperparameter)

Classification Report (Test Set)

	precision	recall	f1-score	support
Negative	0.80	0.71	0.75	2145
Neutral	0.55	0.50	0.53	3000
Positive	0.92	0.95	0.93	15321
accuracy			0.86	20466
macro avg	0.76	0.72	0.74	20466
weighted avg	0.85	0.86	0.85	20466

Findings, Insights, & Analysis:

1) CR and CM show that F&B customers are more likely to leave positive reviews

2) The model works well on positive and negative reviews, but neutral reviews may not be so accurate because of a mismatch between the number of stars given and the tone of language used

Confusion Matrix (Test Set)