Problem: - Multi-class Classification Problem.

Given:- train and test data.

Understanding:- 1. Given train, data have 20k records and 6 classes.

2. Classes are high imbalanced.

3. It is a multi-class classification problem.

4. We have to implement NLP based algorithm which has a good understanding of language.

5. We can use transfer learning.

6. We need an algorithm for weight balancing.

7. Sentences are variable in length so we need robust embeddings.

8. We need a word-piece tokenizer

9. Need Classification report.

Solution 1:- Transformer-based Distill BERT.

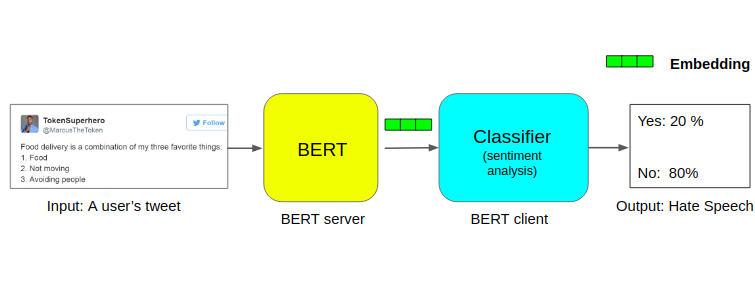
Explanation:- The Transformer is the first transduction model relying entirely on self-attention to compute representations of its input and output without using sequence-aligned RNNs or convolution.

DistilBERT is a small, fast, cheap, and light transformer model based on the BERT architecture. Knowledge distillation is performed during the pre-training phase to reduce the size of a BERT model by 40%.

Result:- Model is highly BIASED. Got 48% accuracy.

Solution 2:- Transformer-based BERT.

Explanation:- BERT Base: 12 layers (transformer blocks), 12 attention heads, and 110 million parameters



\*\*\* To balance the weight I have used the class-imbalance technique.

After balancing the weight I applied to cross-entropy loss.

So that while training neural networks our model can give high weightage to those classes which are less in the count.

Result - Model is low biased but high variance. Got 95% train accuracy but 50% test accuracy.

Hyperparameters: - 0.01 lr, 0.001 lr, epocs= 4, 6, 10

.

Conclusion: - DistillBert model is highly biased. BERT model is low biased and high variance.

NEXT STEP: - we can use XLNET or ROBERT for reducing the variance in our model.