Tools and libraries used :

Tensorflow, numpy, pandas, Scikit-learn

Solution :

1. We concatenated all the textual features (DESCRIPTION BULLET\_POINTS, TITLE) into a single column
2. We removed the BRAND column as it had a lot of noise
3. We mapped the BROWSE\_NODE\_ID column to a sequence of integers (0 to 9919) as the node ids were not consistent
4. We removed the classes whose frequency were less than 500
5. The final dataset had two columns (text features and targets) and 500,000 rows with 863 classes
6. This dataset is split into train and test sets
7. We used a keras Tokenizer class to map each word to a unique integer (maximum integer value : 2499)
8. We padded with the <pad> token for consistency
9. We built a neural network with an embedding layer of dimension 1024, followed by two GRU layers (1024 units each). There is a final Dense layer with softmax activation
10. The model was trained for an epoch and the weights were saved
11. The saved weights were loaded and used for predictions in the test data