1. Tokenisation.py needs to be imported.

pip install sentencepiece

pip install tokenizer

1. Configuration parameter need to be set:

input\_training\_master\_filepath = 'master\_train.csv'

input\_training\_node\_filepath = 'node\_train.csv'

input\_test\_master\_filepath = 'master\_test.csv'

input\_test\_node\_filepath = 'node\_test.csv'

Output\_file\_path = 'result\_submission.csv'

bert\_module\_url = 'https://tfhub.dev/tensorflow/bert\_en\_uncased\_L-12\_H-768\_A-12/2' # default path

‘Alert ID’ column contains binary output- 0 or 1

‘Threat Level’ column contains confidence score

1. Pre-requisite: Update Alert ID field in **(training data)**

A binary value 0 or 1 needs to be populated in ‘Alert ID’ to indicate classification