Natural language preprocessing

Week 1

1. Tokenizer will handle the heavy lifting for us, generating the dictionary ofword encodings and creating vectors out of the sentences. (tensorflow.keras.preprocessing.text.Tokenazer)
2. OOV token - a special token that is for words that aren't recognized, that aren't in the word index itself
3. Then convert the words in those sentences to sequences of tokens by calling the text_to_sequences method.
4. tf.keras.preprocessing.sequence.pad_sequences – pads sequences to the same length creating a matrix of preprocessed sentences

Week 2

Week 3

5. tf.keras.layers.Embedding(vocab_size, embedding_dim, input_length=max_length) – embedding layer keras
6. tf.keras.layers.GRU() – GRu keras layer
7. tf.keras.layers.Bidirectional() – Bidirectional (read the whole sentence from start to the end and vice versa
8. tf.keras.layers.LSTM(64, return_sequences=True) – Lstm layer Keras