## **Table with Hyper and Training Parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PARAMETERS** | **MODEL 1** | **MODEL 2** | **MODEL 3** | **MODEL 4** |
| Word2Vec Embedding Vector Size | NAN | NAN | 50 | 75 |
| Word2Vec Window Size | NAN | NAN | 2 | 5 |
| Number of Encoder Layer | NAN | NAN | 4 | 4 |
| MultiHeadAttention Heads | NAN | NAN | 8 | 2×4=8 |
| Feedforward Network Neurons | NAN | NAN | 128 | 128 |
| Activation Function | softmax | ReLU,softmax | ReLU | ReLU,Sigmoid |
| Learning Rate Initial | 0.001 | 0.001 | 0.05 | 0.001 |
| Learning Decay Rate | NAN | NAN | 0.85 | 0.85 |
| Learning Rate Decay Steps | 0.001 | NAN | 20000 | 20000 |
| Dropout Rate | 0.3 | 0.3 | 0.1 | 0.1 |
| Batch Size | 64 | 64 | 33 | 32 |
| Epochs | 50 | 30 | 15 | 15 |
| Pre-trained bert embedding | 768 | NAN |  |  |

## **Models performance On Evaluation Metrics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **METRICES** | **MODEL 1** | **MODEL 2** | **MODEL 3** | **MODEL 4** |
| BLEU-1 | 4.1979×10−155 | 0.1500 | 0.8781 | 0.9082 |
| BLEU-1 | 4.1979×10−155 | 0.1251 | 0.8781 | 0.9082 |
| BLEU-1 | 4.1979×10−155 | 0.1022 | 0.8792 | 0.9090 |
| BLEU-1 | 4.1979×10−155 | 0.0660 | 0.8780 | 0.9081 |
| ROUGE (F1) | 0.2978 | 0.2466 | 0 | 0.4250 |