```
BertForSequenceClassification(
    (bert): BertModel(
        (embeddings): BertEmbeddings(
            (word_embeddings): Embedding(30522, 768, padding_idx=0)
            (position_embeddings): Embedding(512, 768)
            (token_type_embeddings): Embedding(2, 768)
            (LayerNorm): BertLayerNorm()
            (dropout): Dropout(p=0.1, inplace=False)
        )
        (encoder): BertEncoder(
            (layer): ModuleList(
                (0): BertLayer(
                    (attention): BertAttention(
                        (self): BertSelfAttention(
                            (query): Linear(in_features=768, out_features=768, bias=True)
                            (key): Linear(in_features=768, out_features=768, bias=True)
                            (value): Linear(in_features=768, out_features=768, bias=True)
                            (dropout): Dropout(p=0.1, inplace=False)
                        )
                        (output): BertSelfOutput(
                            (dense): Linear(in_features=768, out_features=768, bias=True)
                            (LayerNorm): BertLayerNorm()
                            (dropout): Dropout(p=0.1, inplace=False)
                        )
                    )
                    (intermediate): BertIntermediate(
                        (dense): Linear(in_features=768, out_features=3072, bias=True)
                    )
                    (output): BertOutput(
                        (dense): Linear(in_features=3072, out_features=768, bias=True)
                        (LayerNorm): BertLayerNorm()
                        (dropout): Dropout(p=0.1, inplace=False)
                    )
                )
            )
        ) … Other 11 layers
        (pooler): BertPooler(
            (dense): Linear(in_features=768, out_features=768, bias=True)
            (activation): Tanh()
        )
    )
    (dropout): Dropout(p=0.1, inplace=False)
    (classifier): Linear(in_features=768, out_features=2, bias=True)
)
```