

Models :

BERT,RoBERTa,DistilBERT,MegaTRONBERT :

Configurations used :

```
"config_1" : {"batch_size" : 8,"epochs" : 20,"learning_rate" : 3e-6,"word_max_len" : 64, "h1" : 768,  
"h2" : 128,"drop_out_rate" : 0.2},
```

```
"config_2" : {  
  "batch_size" : 16,  
  "epochs" : 20,  
  "learning_rate" : 1e-6,  
  "word_max_len" : 64,  
  "h1" : 768,  
  "h2" : 128,  
  "drop_out_rate" : 0.2
```

```
},
```

```
"config_3" : {  
  "batch_size" : 32,  
  "epochs" : 20,  
  "learning_rate" : 2e-6,  
  "word_max_len" : 64,  
  "h1" : 768,  
  "h2" : 128,  
  "drop_out_rate" : 0.2
```

```
},
```

```
"config_4" : {  
  "batch_size" : 64,  
  "epochs" : 20,  
  "learning_rate" : 1e-6,  
  "word_max_len" : 64,
```

```
"h1" : 768,  
"h2" : 128,  
"drop_out_rate" : 0.2  
}
```

Accuracy reached for each configuration:

BERT : 0.98,0.97,0.95,0.97

RoBERTa : 0.96,0.95,0.97,0.95

DistilBERT: 0.96,0.96,0.96,0.95

MegaTRONBERT : 0.95,0.95,0.98,0.95

Time to train the model on GPU P100:

BERT : 2 hr

RoBERTa : 1hr 40 min

DistilBERT: 1hr 50 min

MegaTRONBERT : 2 hr

Datasets chosen

- cyber-bulling-large-dataset
- siber-zorbalk
- cyberbully-detection-dataset
- d/shauryapanpalia/cyberbullying-classification
- cyberbullying-classification
- cyberbullying-dataset
- cyberbullying-tweets