# Methodology

## Base Approach established:


The Team went with the use of ERNIE 2.0 to carry out the task of Sentence Label Classification which has been established to have outperformed BERT and the recent XLNet in 16 NLP tasks in Chinese and English. The base model contains 12 layers, 12 self-attention heads and 768-dimensional of hidden size while the large model contains 24 layers, 16 self-attention heads and 1024-dimensional of hidden size. The model settings of XLNet are the same as BERT. In our approach we decided to move with the Ernie Base model.

## arguably_1:

The method used the above shown pipeline directly since it is robust enough to handle the classification task efficiently. The data before getting loaded to the Ernie Model for processed as mentioned:

**(1) Tokenize the sentence.**

**(2) Prepend the `[CLS]` token to the start.**

**(3) Append the `[SEP]` token to the end.**

**(4) Map tokens to their IDs.**

**(5) Pad or truncate the sentence to `max_length` i.e., 200 in our case**

**(6) Create attention masks for [PAD] tokens.**

The processed embeddings were then fed into the model to attain the desired results.

Batch Size: 16

GPU used for training: **Tesla P100-PCIE-16GB**

## arguably_2:

In addition to the steps mentioned in arguably_1 run the data was pre-processed in this case. The team used text cleaning steps i.e.,

**(1) Stop words Removal**

**(2) Punctuations Removal**

**(3) Lemmatizing**

**(4) Stemming.**

**Also, it was observed that the sentence with length less than 8 mostly acted as the continuation of previous idea and thus the team decreased the size of dataset by appending this smaller sentence to its previous continuation since there was no change in the output labels.**

**For e.g.:**

| | | |
|---|---|---|
| 9 | (ii) orders, rules regulations and bye-laws issued by the State Government under the Constitution of India or under any law made | Statute |
| 10 | (b) all or any of the official purposes of the State; and different dates may be appointed for different purposes in cls. | Statute |
| 11 | (a) and (b) aforesaid. | Statute |

**In above shown image the sentence 11 can be merged with the sentence 10 so as to decrease data size as well as improve robustness of model.**

Besides the above-mentioned text pre-processing same steps were followed as in arguably_1 run to yield final results.

**Some Other tried approaches:**

**1.Use of Law-Bert as the base model however no significant results were observed on the validation set**

**2.Use of Base Roberta as the base model.**