# Tasks Results

**Task 1: Text classification using BERT and comparison with simple LSTM model performance**

**Task 2: Text based clustering (NLP) : Perform unsupervised topic modelling of unlabeled text**

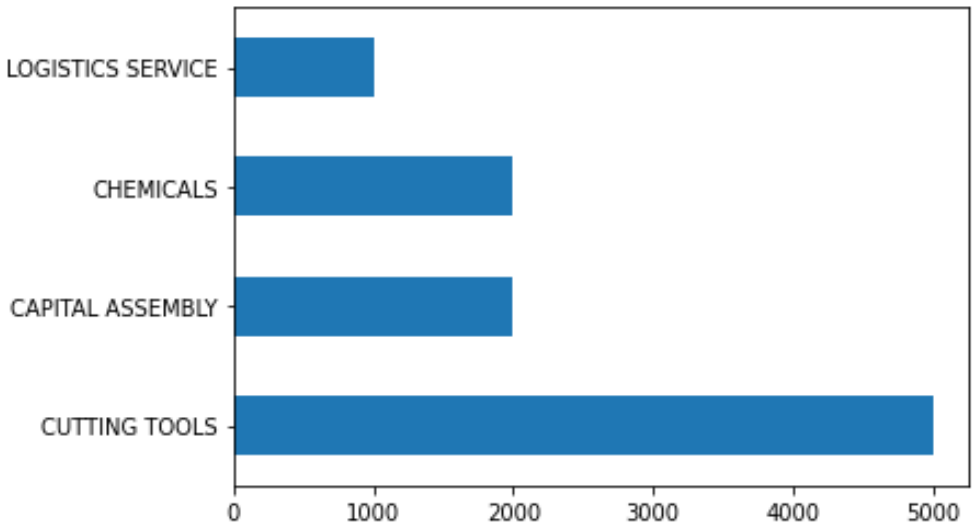
**Task 3: Write a function that takes text array as input and returns the data with corrected spellings.**

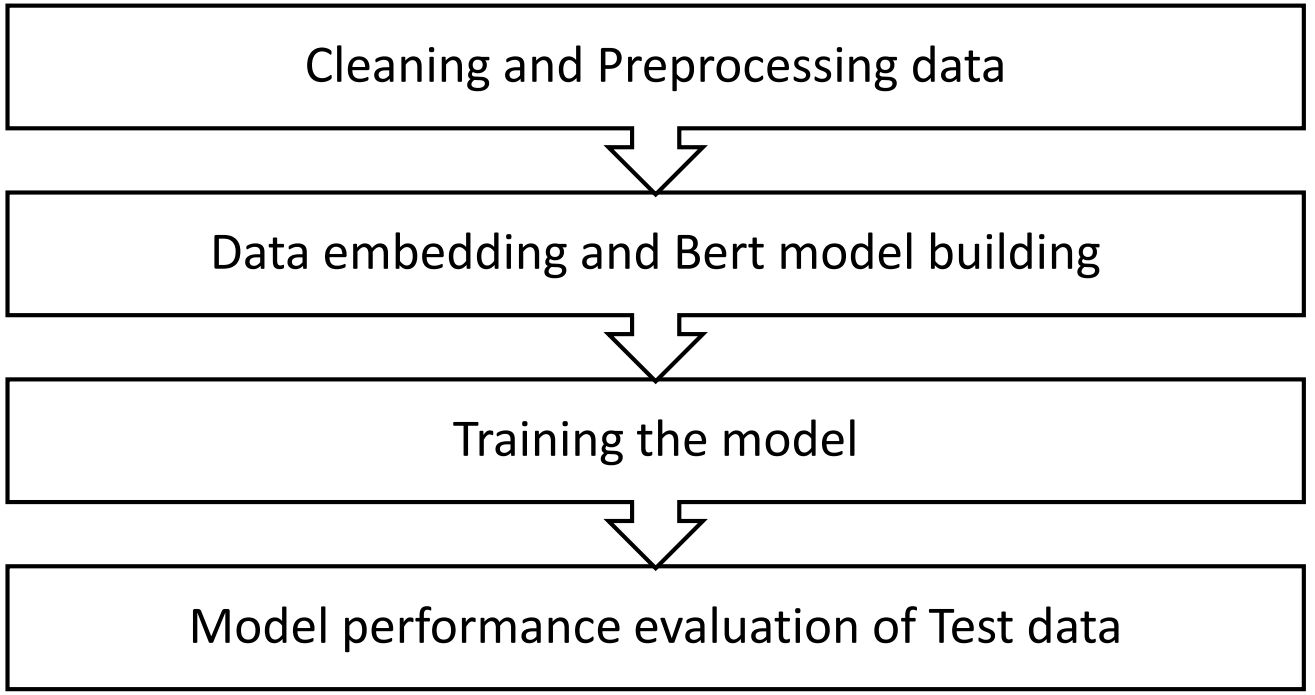
Sachinkumar Revankar
sachinkumarme093@gmail.com
Mob:9008570443

# Task 1: Text classification using BERT and comparison with simple LSTM model performance

**Data detail:**

- The given **Bert_Sample.csv** file has 10000 text data points with 4 classes to classify
- Below is the data points each classes having
- Transformer based pre trained BERT language mode is fine tuned to classify the 4 different classed and the results as follows



# Task 1: Text classification using BERT and comparison with simple LSTM model performance

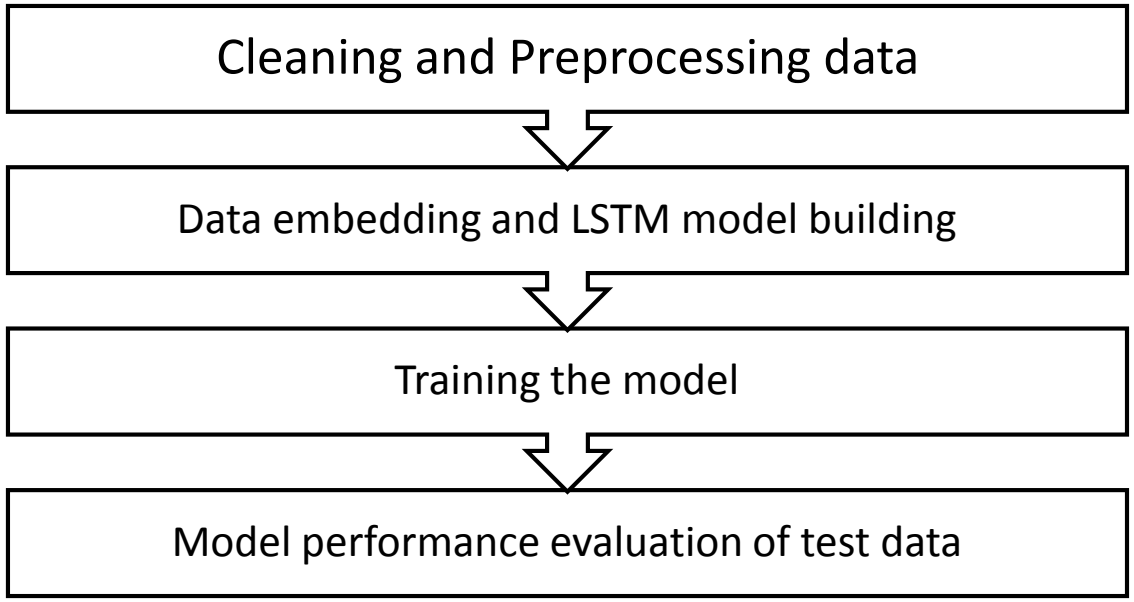
- BERT model results on test data

```
Accuracy of BERT is: 0.928
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| CAPITAL ASSEMBLY | 0.87 | 0.94 | 0.90 | 207 |
| CHEMICALS | 0.92 | 0.81 | 0.86 | 200 |
| CUTTING TOOLS | 0.98 | 0.99 | 0.99 | 504 |
| LOGISTICS SERVICE | 0.78 | 0.84 | 0.81 | 89 |
| accuracy | | | 0.93 | 1000 |
| macro avg | 0.89 | 0.89 | 0.89 | 1000 |
| weighted avg | 0.93 | 0.93 | 0.93 | 1000 |

# Task 1: Text classification using BERT and comparison with simple LSTM model performance

- Steps followed in LSTM model to classify




**LSTM model Results**

```
Test set
  Loss: 0.546
  Accuracy: 0.847
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| CAPITAL ASSEMBLY | 0.85 | 0.76 | 0.80 | 211 |
| CHEMICALS | 0.70 | 0.83 | 0.76 | 177 |
| CUTTING TOOLS | 0.91 | 0.91 | 0.91 | 527 |
| LOGISTICS SERVICE | 0.80 | 0.72 | 0.76 | 85 |
| accuracy | | | 0.85 | 1000 |
| macro avg | 0.82 | 0.80 | 0.81 | 1000 |
| weighted avg | 0.85 | 0.85 | 0.85 | 1000 |

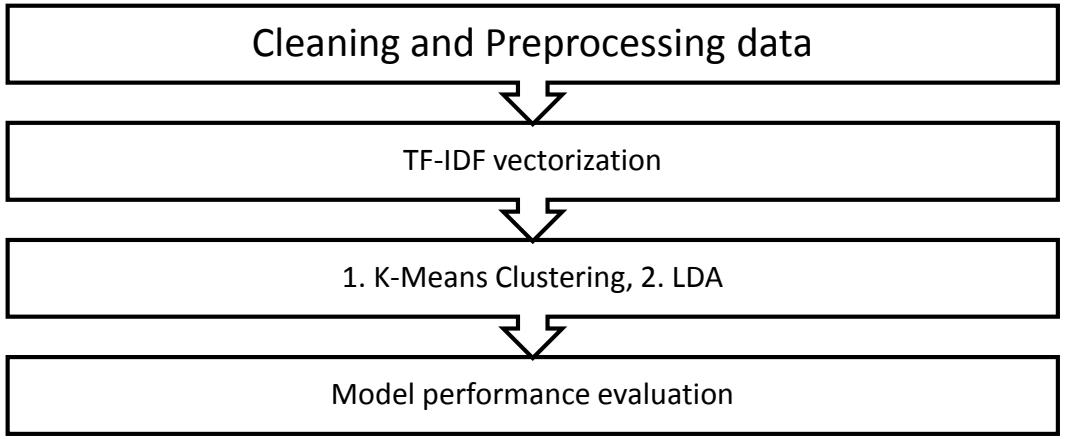
**Conclusion:**

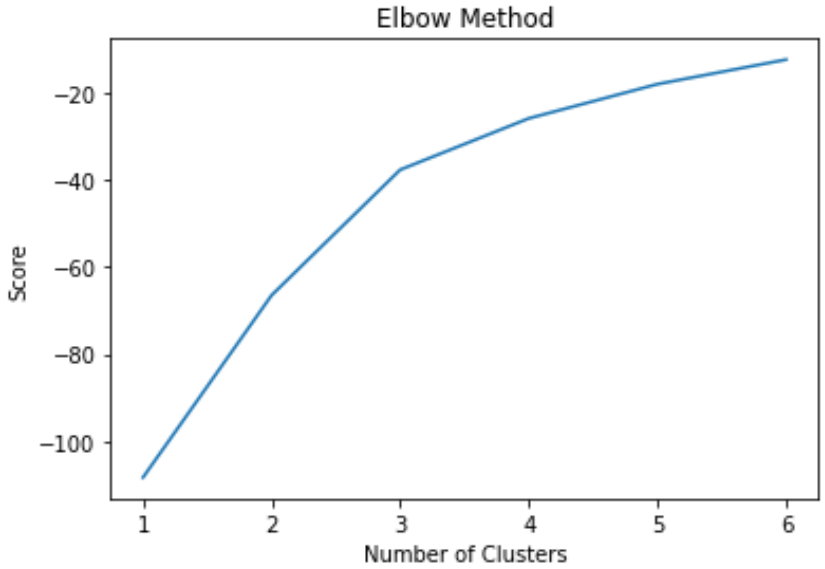
**For given text data, BERT model accuracy is more than LSTM model**

## Task 2: Text based clustering (NLP) : Perform unsupervised topic modelling of unlabeled text

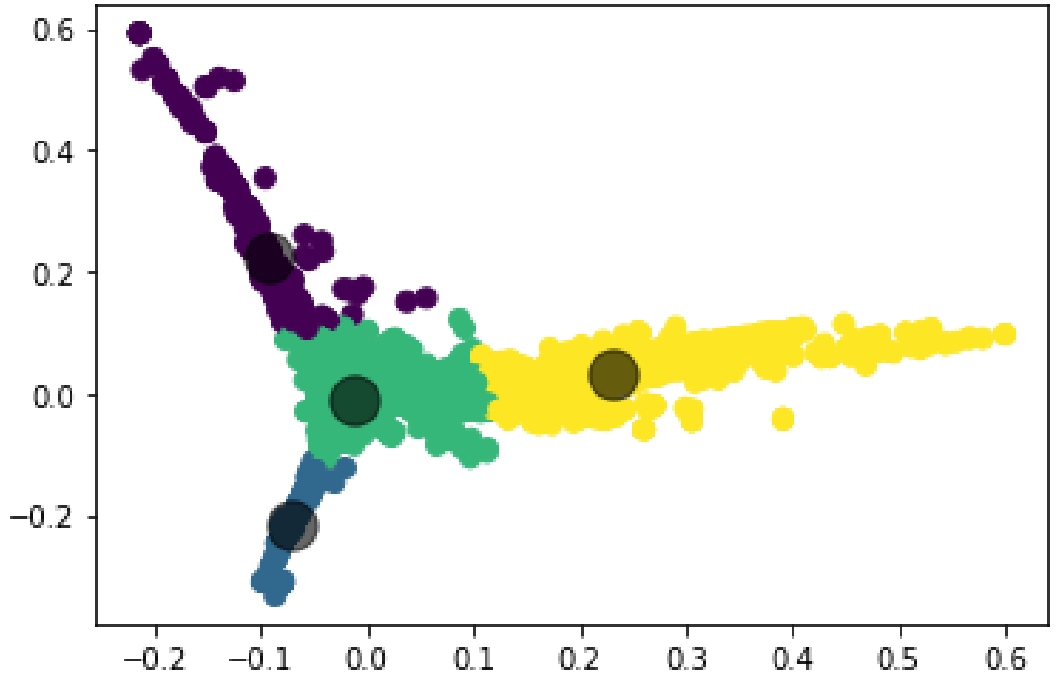
- In present work, text based clustering is made with 2 approaches

1. **TF-IDF and K-Means Clustering**
2. **Latent Drichilet Allocation**




**K-Means Clustering Results:**



**Task 3: Write a function that takes text array as input and returns the data with corrected spellings.**

Spelling correction task is performed by with use of 2 library:

1. **TextBlob**
2. **Gingerit**

# Thank You