# CSCI 544 - Applied Natural Language Processing
# Homework 4
# Report

## Task 1:

Model description:

- Embedding layer
- BLSTM
- Dropout layer
- Linear Layer
- ELU layer
- Classifier layer

Parameters include:
Embedding layer = 100
Hidden dimension = 256
Output dimension = 128
LSTM layers = 1
Dropout value = 0.33
Epochs = 20
Learning rate = 0.1

```
BiLSTM(
  (embedding): Embedding(30292, 100)
  (LSTM): LSTM(100, 256, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=128, bias=True)
  (dropout): Dropout(p=0.33, inplace=False)
  (elu): ELU(alpha=0.01)
  (classifier): Linear(in_features=128, out_features=9, bias=True)
)
```

Evaluation metrics:

| | |
|---|---|
| Precision | 74.63 |
| Recall | 67.80 |
| F1 Score | 71.80 |

## Task 2:

Model description:

- Embedding layer
- BLSTM
- Dropout layer
- Linear Layer

ELU layer
Classifier layer

Parameters include:
Embedding layer = 100
Hidden dimension = 256
Output dimension = 128
LSTM layers = 1
Dropout value = 0.33
Epochs = 10
Learning rate = 0.1

```
BiLSTM(
  (embedding): Embedding(30292, 100)
  (LSTM): LSTM(100, 256, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=128, bias=True)
  (dropout): Dropout(p=0.33, inplace=False)
  (elu): ELU(alpha=0.01)
  (classifier): Linear(in_features=128, out_features=9, bias=True)
)
```

As NER task is case sensitive, to take care of this problem, in the embedding matrix of the word is present then we create its embedding or else we check if its lower word is present, if even that is not present then we create its embedding using the 'UNK'. It's the GloVe embedding, after reading it, I have added two more vector embeddings to it, which are PAD and UNK

```
glove_emb["<PAD>"] = np.zeros((100,), dtype="float64")
glove_emb["<UNK>"] = np.mean(glove_vec, axis=0, keepdims=True).reshape(100, )
```

Evaluation metrics:

| | |
|---|---|
| Precision | 87.11 |
| Recall | 88.71 |
| F1 Score | 87.90 |