**Machine Learning 3 Assignment**

**Unsupervised Learning**

**Assignment Team**

* **Subbalakshmi Narayan Prasad (Pre-processing) 21250754**
* **Naveen Ramprasad (Data retrieval and Model development) 20253139**

**Objective:**

To find meaningful structure from the text files input data given using any unsupervised algorithm.

**Chosen idea**: Use the trained model to generate meaningful goals, given the words to start with.

**Methodology:**

1. Read the input data and extract the Goals from the text files.
2. Pre-process the text files data
3. Develop a basic sequential model with LSTM to train
4. Input the starting words from the user to generate a paragraph of 200 words

**Pre-Processing:**

1. Extract necessary data (Goals) from the text files. Some documents have mentioned the heading as ====Goals and others as == Goals. Both of these ways are accepted in the pre-processing
2. Remove punctuations and decode it to ascii
3. Tokenize the data
4. Add padding to ensure that the sentences are of same length

**Model Building:**

1. Develop a basic sequential model with embedding and one hidden layer
2. Add LSTM hidden layer
3. Add a softmax classifier for the output layer to get the classes in probabilities
4. Use adam optimizer to compile the gradient descent. The loss function used is cross entropy
5. Fit the model for 100 epoch
6. Predict the sentence completion by giving the start of a sentence of your wish

**Result:**

Given the start of the sequence, we have successfully predicted a goal with very few punctuation errors that is comparable to the goal shown in the training data.