**Lab Assignment 2**

**Maninandan Kumar**

**Id: 16233134**

**Introduction:**

CNN is a machine learning model used for visual imagery and it is designed to have minimal preprocessing through multi layer perception. Earlier CNN was applied to images but now it is also being applied to text data. In this task, we are applying CNN model on given dataset and process text classification.

**Objectives:**

Apply CNN model on dataset and obtain text classification results.

**Approches:**

1. Load data from files byclassifying it positive and negative
2. Clean the text data
3. Pad each sentence to the maximum sentence length. Padding sentences to the same length is useful because it allows us to efficiently batch our data since each example in a batch must be of the same length.
4. Build a vocabulary index and map each word to an integer between 0 and 18,765 (the vocabulary size). Each sentence becomes a vector of integers

**Parameters:**

Input\_x= sequence\_length

Input\_y = num\_classes

vocab\_size

**Datasets:**

In this assignment we have taken dataset consisting of crime records database and classify the text into positive and negative.

**Conclusion:**

Because dev accuracy is significantly below training accuracy it seems like our network is overfitting the training data, suggesting that we need more data. The training loss and accuracy starts out significantly below the dev metrics due to dropout applied to it.