**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:**

Load the data from file, clean and train data and testing data and getting results

Train data Command: python3 train.py ./data/train.csv.zip ./training\_config.json

While training the data a directory is created and trained data is saved in a new directory

This new directory is used as model for predicting against the new data set

Predicting Command: python3 predict.py/trained\_results /new\_data\_file

**Parameters:**

Input\_x= sequence\_length

Input\_y = num\_classes

vocab\_size, sequence\_length, num\_classes, filter\_sizes, and num\_filters

**Datasets:**

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

**Conclusion:**

CNN classification on dataset is implemented and dataset is classified, 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.