**Deep Learning**

**Satya Harish Gumpalli**

**Class ID: 13**

**Student ID: 16231349**

**Lab Assignment-2**

**Text Classification with CNN**

Introduction:

Text Classification is one of the critical parts of deep learning. As most of the people communicate through text in the form of tweets, blogs, messages, emails, and comments. Now a days, NLP techniques are dealing with this by using words as symbols and running linear models. It is used to achieve good classification among range of classification tasks. With Convolution we apply a moving window to your input data and letting the neural network learn the weights to apply to adjacent words.

Objectives:

Here, the objective is to implement the text classification with CNN model. Using this model, the category of the crime can be predicted. For this assignment, I have used the dataset Kaggle San fransisco Crime.

Approach/Method:

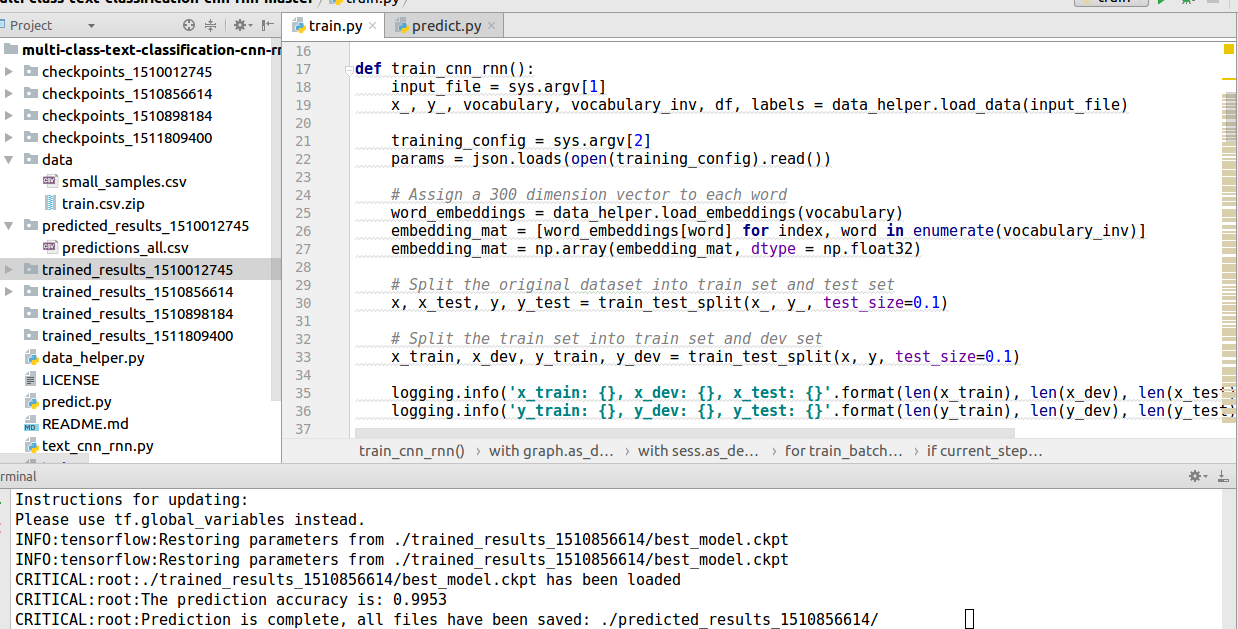
In this approach, I have implemented CNN network by using dataset Kaggle San Fransisco Crime. This is a multiclass text classification problem, which consists of 39 classes. Using the text classification with CNMN model, we are going to classify the crime into 39 classes. This dataset contains crime incidents recorded from SFPD Crime incident reporting system. It contains the data fields such as Dates, Category, Descript, Day of week, PdDistrict, Resolution and address. Here, I have developed a model of multiple layers. First layer embeds words into low-dimensional vectors. The next layer performs convolutions over the embedded word vectors using multiple filter sizes.

Workflow:

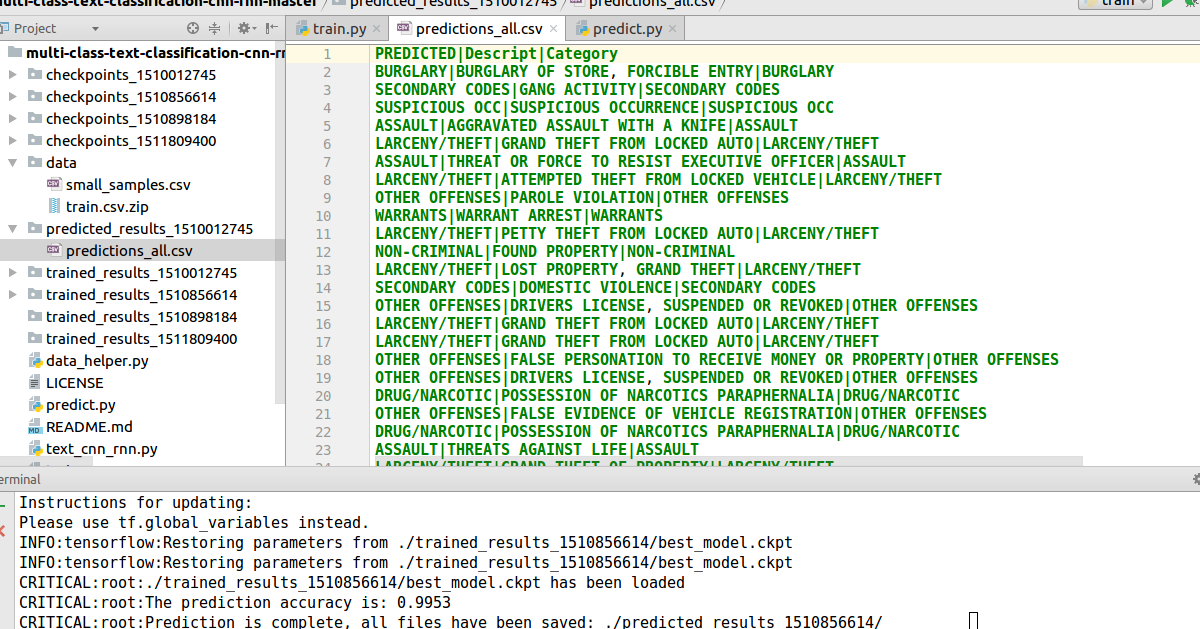
I did various research on text classification libraries and finally I decided to go with CNN to solve this problem. Different steps that I have followed are getting the data, training the data, testing the data and getting the results. I have used python script to scrape all the categories and then I got the data. The parameters that I used are sequence\_length, num\_classes, vocab\_size, embedding\_size, filter\_sizes, and num\_filters. Once the data was ready, I have trained the data using the command python3 train.py ./data/train.csv.zip ./training\_config.json . Then I did run the code on the trained data using the command python3 predict.py ./trained\_results\_1478563595/ ./data/small\_samples.csv and tensorflow created a new folder called trained\_results, which contain the final results. Then after we will visualize the graphs on the tensorboard.

Evaluation/Discussion:

This is the code train.py



This is the dataset file predictions\_all.csv



trained\_results\_1510856614 file that is generated when we use the command python3 predict.py ./trained\_results\_dir/ new\_data.csv. Here, category of the crime is predicted.

