Manual

This manual contains instructions and tips for running each of the three pertinent .py files.

cnn_classification.py

This file contains the CNN model which was developed in the report. To run it, simply run the Python file using whichever method you wish.

The dataset used in training and evaluation can be modified on line 52, found within the 'Load & Preprocess Data' section. Below is a screenshot of the code snippet for reference.

```
# ==================== 2. Load & Preprocess Data ====================


# Load the dataset
df = pd.read_csv("tensorflow.csv")
```

There are five available datasets to pick from:

tensorflow.csv

pytorch.csv

keras.csv

incubator-mxnet.csv

caffe.csv

The 'NUM_RUNS' variable on line 173 dictates how many times the training and testing loop is carried out. This is currently set to five, but can be modified to your liking.

```
173     NUM_RUNS = 5
```

br_classification.py

This file contains the baseline TF-IDF and Naïve Bayes model. To run it, simply run the Python file using whichever method you wish.

cnn_optimiser.py

This file contains the code which was used in the hyperparameter tuning process. To run it, simply run the Python file using whichever method you wish.

The 'n_trials' field on line 216 dictates the number of hyperparameter configurations that will be tested. This is currently set to 100, but can be modified to your liking. If you only want to see

a flavour of how the code works, it is recommended to reduce this number, as it takes several hours to complete 100 tests.

```
216        study.optimize(objective, n_trials=100)
```

As with cnn_optimisation.py, the NUM_RUNS variable dictates how many times a training and testing loop is done for the given model. In cnn_optimiser.py, this can be found on line 106.

```
105        # Run each configuration 5 times and average the metrics
106        NUM_RUNS = 5
```