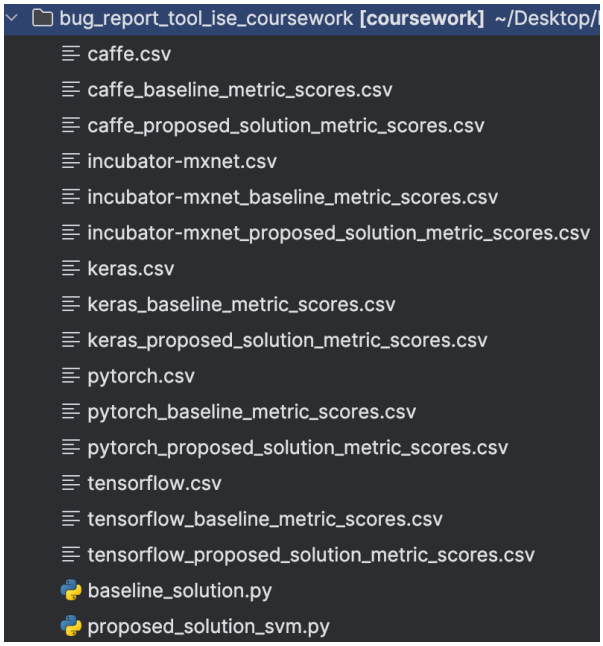
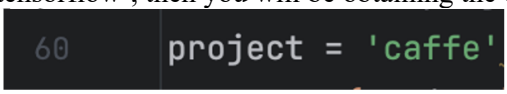
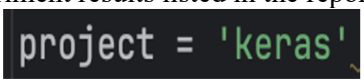
Using the Proposed Solution Tool

Step 1: Download **the bug_report_tool_ise_coursework folder in the GitHub Repository**
Step 2: Please ensure that the datasets (**caffe.csv, incubator-mxnet.csv, keras.csv, pytorch.csv, tensorflow.csv**), **proposed_solution_svm.py** and **baseline_solution.py** are in the same directory as shown below: The folder will also include the metric_scores.csv which are the raw results.



Step 3: In our tool, proposed_solution_svm.py at line 60, we have set a default dataset, 'caffe'. Please do change this string use our tool to obtain the reported experiment results on that dataset, the available datasets are in the folder you have downloaded and are csv files. The screenshots below show what strings can be stored in the project variable . E.g. if you change line 60, to indicate project = 'tensorflow', then you will be obtaining the experiment results listed in the report for the tensorflow





E.g. If you wish to run our tool using the 'caffe.csv' dataset, then please leave it as it is and proceed to step 4. Otherwise, please change the string held in the project variable to hold the name of the desired dataset you wish to run with our tool before proceeding to step 4. You will be required to change this string to obtain the results reported in the report based on which results you would like to reproduce and verify.

Please ensure that the specified dataset csv file (i.e. dataset) is in the same directory as the , **proposed_solution_svm.py** file before executing.

Step 4: To run our tool, please execute the following command in the same directory as the proposed_solution.py file and the csv file stated in line 61:

```
python proposed_solution_svm.py
```

Note: try 'python3 proposed_solution_svm.py' if the above does not work.

Step 5: The tool repeats the process of fitting an SVM model 10 times for robustness. For each run, we print out on the console the evaluation metric scores and optimal parameters . E.g. Below shows the evaluation metric scores on the 5th iteration whilst using the caffe.csv dataset for training and testing:

```
Optimal Parameter Setting for Repeated Experiment Run 5: {'C': 10, 'gamma': 0.01, 'kernel': 'rbf'}
Repeated Experiment Run 5: Accuracy:      0.7759
Repeated Experiment Run 5: Precision:      0.6101
Repeated Experiment Run 5: Recall:      0.7276
Repeated Experiment Run 5: F1_Score:    0.6221
Repeated Experiment Run 5: AUC:    0.7276
```

Please note that optimal parameter settings varied across the 10 runs due to the randomness of the splitting the dataset into a training and testing set. We have documented this in our report as figure 5.

Below shows the mean results across the 10 runs for each evaluation metric for our proposed solution. These solutions were documented in our report as figure 3,4,5,7 and were compared with the mean results across 10 runs for each evaluation metric for the baseline solution.

```
=== SVM + TF-IDF Results ===
Number of repeats:      10
Average Accuracy:       0.8017
Average Precision:      0.5990
Average Recall:         0.6350
Average F1 score:       0.5968
Average AUC:            0.6350
```

In addition to this, in the console we also print statistical test results between the proposed solution and baseline solution. Below is a screenshot of what you would expect to see after executing the proposed solution code using the caffe dataset. We compared the baseline precision values array with the proposed solution values array. This was repeated using Recall and F1 values from the baseline and proposed solution

```
=== Wilcoxon Rank Sum Test Results ===
Wilcoxon Rank Sum P-Value for Precision 0.19876460637323512
Wilcoxon Rank Sum P-Value for Recall 0.6501474440948545
Wilcoxon Rank Sum P-Value for F1-Score 0.010165201891956 26
```

We also print out the standard deviation of the 10 values from the repeated experiment for each evaluation metric. This can then be verified with the figures shown in the report in figures 3,4,5,6,7

```
=== Standard Deviation Results ===
Standard Deviation for Accuracy:      0.0464
Standard Deviation for Precision:     0.0658
Standard Deviation for Recall:        0.0800
Standard Deviation for F1 score:      0.0655
Standard Deviation for AUC:           0.0800
```

The all the above results printed in the console are stored in a csv filed namely, **{project_name}_proposed_solution_metric _scores.csv**, where project name is the name of the dataset that has been specified in line 60 of the baseline solution,

Using the baseline code

We have included the baseline code, which is called, baseline_solution.py in our GitHub repository as we have made modifications to the one provided by the module team. This can be downloaded the same way you did for the proposed solution as it is in the same directory as the proposed solution and the relevant datasets. The name of the file is **baseline_solution.py** . Please follow step 1 and step 2 as described above.

Step 3: For the baseline solution named, baseline_solution.py at line 68, we have set a default dataset, 'caffe'. Please change this string to one of the following to obtain the reported experiment results on a specified dataset, the available datasets are in the folder you have downloaded and are csv files. The screenshots below show what strings can be stored in the project variable. E.g. if you change line 68, to indicate project = 'tensorflow', then you will be obtaining the experiment results listed in the report for the tensorflow

```
68  project = 'caffe'
```

```
project = 'tensorflow'
```

```
project = 'pytorch'
```

```
project = 'keras'
```

```
project = 'incubator-mxnet'
```

E.g. If you wish to run baseline tool using the 'caffe.csv' dataset, then please leave it as it is and proceed to step 4. Otherwise, please change the string held in the project variable to hold the name of the desired dataset you wish to run with our tool before proceeding to step 4. You will be required to change this string to obtain the results reported in the report based on which results you would like to reproduce and verify.

Please ensure that the specified project(i.e. dataset) csv file is in the same directory as the , **baseline_solution.py** file before executing.

```
69  path = f'{project}.csv'
```

Step 4: To run the baseline tool, please execute the following command in the same directory as the baseline_solution.py file and the csv file stated in line 69 :

```
python baseline_solution.py
```

Note: try 'python3 baseline_solution.py' if the above does not work.

Step 5: The baseline tool repeats the process of fitting a Naïve Bayes model 10 times for robustness as normal. For each run, we print out on the console the evaluation metric scores. E.g. Below shows the evaluation metric scores on the 5th iteration whilst using the caffe.csv dataset for training and testing the model: The array containing these values over the 10 runs are stored in a csv filed namely, **{project_name}_baseline_metric _scores.csv**, where project name is the name of the dataset that has been specified in line 68 of the baseline solution, these files can also be found in the root directory of the GitHub repo. In addition to this, we have also stated these values in the report as well for further analysis – they can be found in figures 10,11,12,13,14

```
Repeated Experiment Run 5: Accuracy:      0.4655
Repeated Experiment Run 5: Precision:      0.5497
Repeated Experiment Run 5: Recall:      0.6282
Repeated Experiment Run 5: F1_Score:    0.4153
Repeated Experiment Run 5: AUC:    0.6282
```

Below shows the mean results across the 10 runs for each evaluation metric for our proposed solution. These solutions were documented in our report as figure 3,4,5,7 and were compared with the mean results across 10 runs for each evaluation metric for the proposed solution.

```
=== Naive Bayes + TF-IDF Results ===
Number of repeats:     10
Average Accuracy:      0.6000
Average Precision:     0.5608
Average Recall:        0.6532
Average F1 score:      0.4968
Average AUC:           0.6532
```

We also print out the standard deviations of the 10 values from the repeated experiment for each evaluation metric. This can then be verified with the figures shown in the report in figures 3,4,5,6,7 and with **{project_name}_baseline_metric _scores.csv**

```
=== Standard Deviation Results ===
Standard Deviation for Accuracy:      0.0854
Standard Deviation for Precision:     0.0346
Standard Deviation for Recall:        0.0868
Standard Deviation for F1 score:      0.0692
Standard Deviation for AUC:           0.0868
```