Instructions to Use this Tool

1 – Download the Solution.ipynb file, and open it in a Juptyer Notebook
2 – Make sure all the Dependencies & Imports are downloaded in the environment
3 – Recommended to make sure you are using GPU (MPS if on Apple Mac M1 Series) instead of CPU, if not, you may need to check your version of Pytorch to be correct. If You cannot get this to work, that's fine it can still use CPU, it will just be slower.
4 – Download the Datasets Folder and add to the root directory (Same place as your Juptyer Notebook Solution File).
5 – Do Step 4 again for Preprocessed_BERT & Preprocessed_Baseline
6 – Now please run the singular Juptyer Notebook cell and results will start to be produced.
7 – If your file paths are incorrect as you did not download correctly or manually did it, you can edit your file paths at the bottom of the Juptyer Notebook cell. (3 different paths).
8 – If your results are taking too long to be produced and you are running into issues, please reduce the NUM_RUNS variable (Same location as the data paths, at the bottom of the cell), from 30 to a smaller number like 5.

Your Results will be in your root folder. You will have the following:

- Experiment Results Table - Data for each individual run (180 rows)
- AllTables - Average Results for each Dataset/Combined for both models (12 rows)
- Overall Table - Average Results for both models (2 rows)
- BoxPlot.png - F1 Score Box plot for each model & Dataset/Combined, showing variance across the 30 runs (12 Box Plots)
- Statistical_testing_results table - F1 Score Statistical Tests, Wilcoxon value & Paired t-Test value for each dataset/combined and each model alongside the averages for all 6 values (7 rows)