This repository is a solution for Grab’s AI for S.E.A. Safety challenge. Details and dataset can be found at: <https://www.aiforsea.com/safety>

**Descriptions for the files in this repository:**

* safety.py: Contains the main algorithm. The algorithm processes and extracts features from the training data, does a 80-20 train-validation split, trains the model and saves the weights of the best model to the file ‘best\_model.hdf5’.
* test.py: To use with the test set. This file loads the test set and does the appropriate data processing. Afterward, it loads the model and its weights in the file ‘best\_model.hdf5’. Finally, it predicts and displays the ROC AUC score for the test set. Independent of ‘safety.py’.
* best\_model.hdf5 : Saved weights of the training model.
* References.doc: Literatures that I have referred in doing this challenge. Due to time constraint, I will only include the URLs for now.
* Journal.doc: A short summary of my thought process and the steps I did in doing this challenge. Also includes some of the shortcomings for this solution and possible improvements.
* features, labels, test\_features, test\_labels: Folders to copy the respective datasets in.

**Instruction to run:**

* Copy the test features and the test label files into the ‘test\_features’ and ‘test\_label’ folders, respectively.
* To run the main script, copy the training features and training test labels files into the ‘features’ and ‘label’ folders, respectively.
* Run ‘safety.py’ for the main algorithm, ‘test.py’ to check the algorithm with the test set. Note that if you run the ‘safety.py’ file first, it might override the ‘best\_model.hdf5’ file I trained.

**Note:**

The training neural network model can sometimes raise the error ‘ValueError: Only one class present in y\_true. ROC AUC score is not defined in that case’. I did some research online about this error, and it seems to be a bug of the ‘roc\_auc\_score’ function from ‘sklearn.metrics’, and I can’t find a convenient fix for this particular implementation. However, if the error happens, one can run the ‘model.fit’ command again, and training progress won’t be affected.