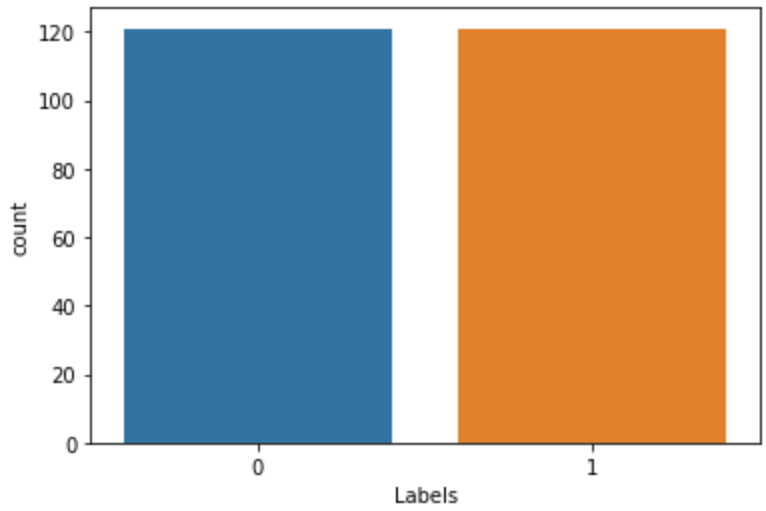
MLBA Assignment 2

Submitted By:
Gaurav Lodhi (MT19063)

Library requirement: Sklearn, Keras, Python 3
Please run as: python3 A2_group21.py kaggle_train.csv kaggle_test.csv result.csv

There is some information how i build the models.

Preprocessing: By using pandas library I read the given train and test CSV files.
I found the following dataset as a train:-



We can see that their classes are balanced so there is no need for sampling. But there is still a need for feature selection because the given dataset is a high dimensional dataset. I used selectKBest and PCA for feature selection.

Methodology: I used sklearn's MLP neural network, xgboost, random forest model to predict the labels and among them, I found that MLP is performing better than other models with the following parameters.

```
random_state=0,hidden_layer_sizes=90, max_iter=35
```

Result:- I got .64 accuracy on the public leaderboard by using the above parameters.