HackerEarth Machine Learning Challenge : Adopt a buddy

The training dataset was imported to Google Colab platform to be used. The ‘pet\_id’ column was excluded from the train data. Upon Analysing it was observed that ‘condition’ column column contained 1477 ‘NaN’ values. The ‘NaN’ values were replaced with an integer ‘-1’ (also in the test dataset). A new column called ‘total\_days’ was added which contained the total number of days between listing\_date and issue\_date. The values “breed\_category” and “pet\_category” were stored in y and removed from training dataset. The rest of the train data was stored in “x”. Color\_type category contained categorical values, thus Label Encoding was done to that column. Then the whole train dataset was scaled to standardize the independent features present in the data in a fixed range. Then training data was split into 80% train data and 20% test data. Random Forest Classifier was used to predict the output of the rest 20%. Finally , the model was used on the given test data.