**HackerEarth Machine Learning Challenge: Adopt a buddy**

Approach: The main approach was to perform few feature engineering steps and apply most commonly used Machine Learning methods with hyperparameter tuning to improve the overall accuracy in predicting the two target variables.

Feature Engineering: Feature engineering was a very important step in improving the overall model metrics. Following were the feature engineering steps performed:

1. Missing value imputation: The train dataset had condition column which had a large number of missing values. It is very important to fill in these missing values by replacing it with the most frequent value from the condition column using mode imputation or by using KNN imputation. I had used KNN imputation from the Sklearn metric to deal with the missing value in my solution.
2. Color column encoding: The color\_type column had multiple category which needs to be handled before applying Machine Learning models. This categorical column needs to be encoded for which I had used Label encoding approach to do so.
3. Handling dates: The dataset also had issue\_date and listing\_date on which I had performed a data difference to calculate the number of days between the issue\_date and listing\_date. Finally, had considered this new column day\_difference as input to my model and had dropped issue\_date and listing\_date.

Finally, performing a split on the dataset and applying multiple models such as Gradient Boosting Classifier, XGBoost, CatBoost and LGBM models with hyper parameter tuning where I had increased the number of estimators to 1000 to improve the final prediction of target variables.

Tools used: Anaconda Jupyter Notebook