Design Approach Document

The below document details the approach taken to solve the problem of predicting the rating of mutual funds.

**Design decision 1:** **Split the dataset**

By the nature of the mutual fund industry, the handling of funds will vary depending on the type of funds & the risk factors involved. The debt-based funds will be of less risk while the equity based funds will be of higher market risk. Hence the ratings will be based on the fund type / category. Hence the 1st design decision is the split the entire dataset based on the ‘Investment Class’ [Value based investment, Growth & Blended]. When the investment class is empty (Null), create a separate class called ‘Unknown’.

**Design decision 2** **: Imputations**

The data imputations for the missing data will be done within the data subset based on investment class. The data imputations will be done after the split of data subsets based on investment class. This will be due to the fact that the items under any class will be treated separately & hence it will be apt to do mean based imputations within a investment class.

**Design decision 3** **: Different algorithms for different investment classes**

Since the datasets will be separated based on investment class, each subset will be split into train & test. For each data subset, the train data will again be split into X\_train, y\_train, X\_test, y\_test (this is equivalent of train & validation data). Multiple algorithms will be run on this train & validation data. Also for each model, we will use a RandomSearch Cross validation to get the optimal hyper parameters which give best results. Final model will be chosen based on the best results among the different models (with hyper parameter tuned). Hence the final result will be a combination of models – different model for individual data subset – which will be merged to form the final result.