Kaggle Competition Approach

Name: Shivam Negi (23M1508)

1. I have applied XGBoost stands for "Extreme Gradient Boosting” algorithm for **my final submissions.** After that I used my previous submissions to predict the missing values of Main class and Subclass of the current submission file. I used previous submissions in two ways:
   1. Taking predictions of my previous submissions directly.
   2. Predict missing classes through KNN(k nearest neighbours) and replace missing class values from classes of most nearest neighbour.
2. Run code file for **XGBoost.ipynb**
3. Run code file **Again submission.ipynb** :
   1. In this file, I take my **previous submissions** having accuracy greater than 80%, and arrange them from highest accuracy to lowest accuracy. Then search for missing values from ranked submissions files and set those missing values in current submission. And I make submission files again and again to reduce missing values and save this file and submit.
4. Run code **KNN Submission.ipynb** :
   1. In this I fit KNN algorithm on train data and find nearest point for missing class and predict classes through three possibilities:
      1. **When Main class and Subclass both missing** then replace directly through the nearest class of that missing point from train data.
      2. **When Only Main class is missing and Subclass is Present** then set main class according to their sub classes.
      3. **When Only subclasses are missing and the Main class is present** then find the nearest point of missing classes point in train data and then take classes from train data and set them according to the main class present.

5) **My previous submissions** (having accuracy 80% to 83%) are based on Logistic Regression with L1 penalty, Stochastic Gradient descent algo with L1 loss and AdaBoost classifier with logistic regression with L1.