**Hello My name is Mayank Rastogi**

**HackerEarth Machine Learning Challenge Resolving citizens’ grievances** as in this challenge I have using Two models **one is Light GBM and 2nd One is XGBoost** .

Soo My approach is very simple that first.

I have importing necessary libraries from data manipulation and processing .

Now, I have read the data and convert into data frame using pandas pd

Now, **I make a new,columns name label to classify the train and test data so use of this because I have performed feature engineering task by combining train and test data, and if we need test data ,I  have compare to label by creating query in which if statement is, performing operartion on test data where as whole data  combiningly to perform Feature Engineering task and reason I have assigned test labels importance to -1 because on this I have performed predictions.**

Now, I have performing **feature engineering task**  on different columns as such by combining all no of issues , by converting text into lowercase letters , by generalizing countries, by removing constant values, by removing not use full features by featurize columns and featurize data columns into separate different decision dates after all that preprocessing by encoding the label by one hot encoded to convert it into numeric discrete values and lastly I have perform filling null values if any.

**As you can see when in My notebook I, have performing feature engineering task then no changes shown to you**

**when we call df.head() the reason is that we have not combining this toany particular dataframe soo next step is that the task we perfrom above is time to convert it into one dataframe soo its reflects and**

**clearly seen.**

Now, building our model by using query to take the train label and initialize it into new variable to help performing training our data task in our model and also splitting the data\_train into train\_test\_split after splitting,

I have train my model using Light GBM and perform Cross validation gives an accuracy on training data i.e , **🡨>**  **87.07406364404393** .

For, 2nd Model I have train our model on model named as XG Boost classifier that give me an accuracy of **🡨>** **50.971557307800616.**

After that I used the LGBM model i.e xgbm as per my notebook for testing the data because it gives us best accuracy for predicting results for test data,

After that I have created the data frame containing two columns app no and predict the importance which my model is predicted.

Last Part is to create the data frame for submission (**grievances\_submission.csv** i,e; our predicted results) in the form of their app id and their Predicted importance of that particular id.

**Tools I have used is only Google Collab to perform complete task.**