

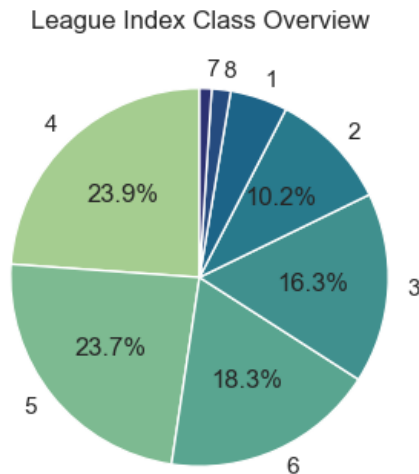
# StarCraft Player Performance – Findings

1. The dataset constituted of performance data of various players in the Starcraft game. We are tasked with predicting the rankings of the players.

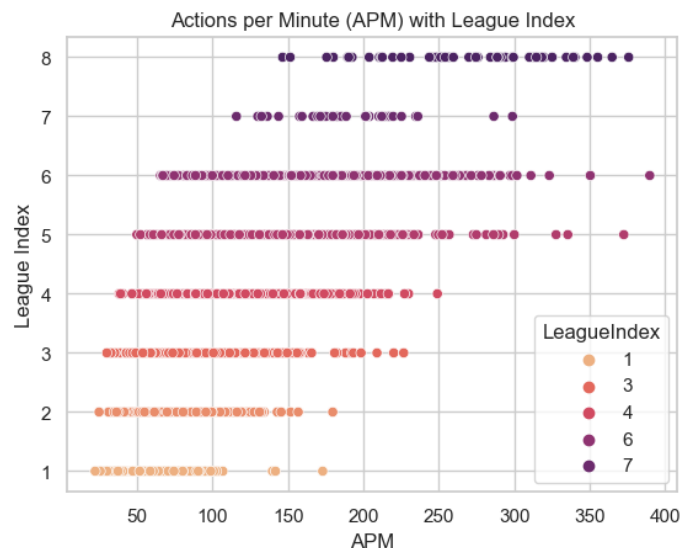
2. A brief description of the dataset includes-

- Player data entries = 3395
- Age of Player
- League Index (Ranking)
- Hours per week and Total hours given to the game
- Actions Per Minute, which tells the number of game actions the player is taking per minute
- Various other features based on gameplay

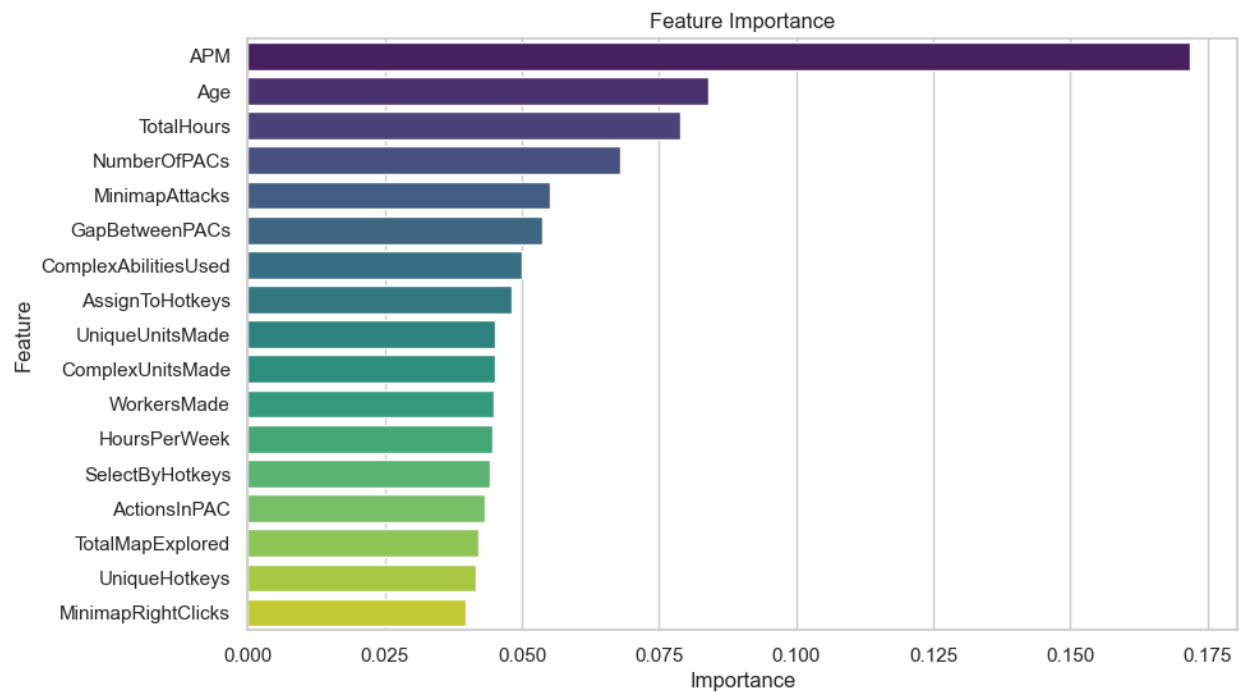
3. The goal was to predict the League Index of the player, which makes it a “Multi-class Classification Problem”. Distribution of Data based on League Index-



4. Through the analysis, I found out that APM (Actions per Minute) had a very high positive correlation with the League Index Rankings.



5. Based on the features we have for the dataset, I tried to find out the most important features for predicting the league Index ranking. These are the results-



6. The dataset was highly imbalanced, which means that the distribution of data based on League Index rankings was not equal. To deal with this, I used two techniques-

a. SMOTE (Synthetic Minority Over-sampling Technique) Oversampling: It creates synthetic samples by interpolating the feature space of minority class instances to generate a more balanced dataset.

b. Evaluation Metric as Balanced Log Loss Function:

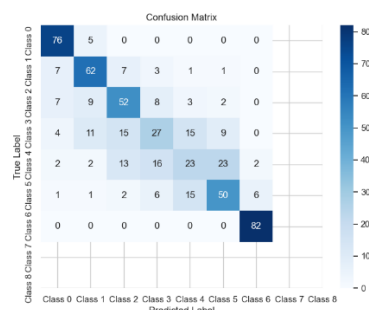
$$Balanced\ Log\ Loss = - \frac{\sum_{i=1}^8 [w_i \sum_{k=1}^{N_m} y_{ik} \log(p_{ik})]}{\sum_{i=1}^8 w_i}$$

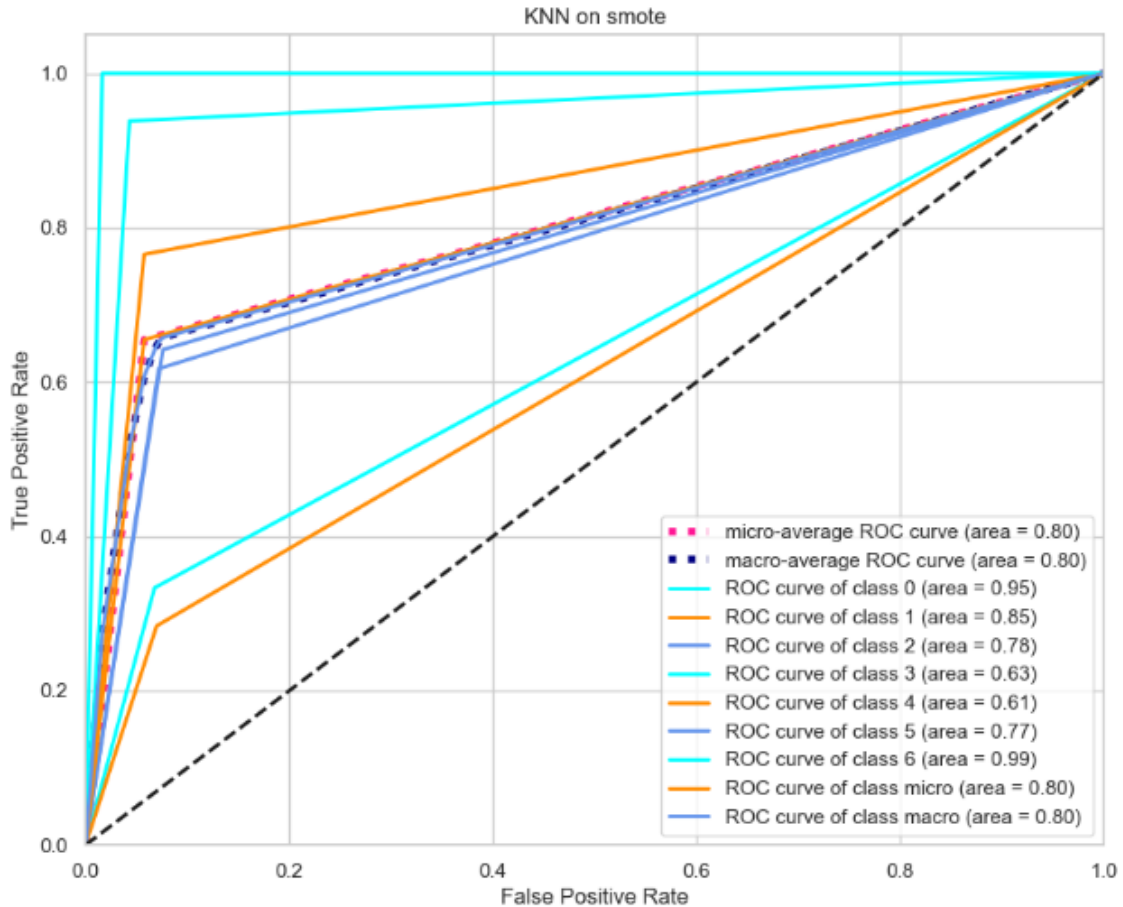
$$where\ w_i = \frac{1}{\frac{Size\ of\ target\ class\ i}{Size\ of\ target}}$$

It basically gives higher weights to the minority classes, making them equally important.

7. I chose K-Nearest Neighbors as the final model to make predictions. These are the results-

	precision	recall	f1-score	support
0	0.78	0.94	0.85	81
1	0.69	0.77	0.73	81
2	0.58	0.64	0.61	81
3	0.45	0.33	0.38	81
4	0.40	0.28	0.33	81
5	0.59	0.62	0.60	81
6	0.91	1.00	0.95	82
accuracy			0.65	568
macro avg	0.63	0.65	0.64	568
weighted avg	0.63	0.65	0.64	568





The League index classes that had good amount of data achieved very high recall and F1 scores. For League index 1 and League index 7, the average recall score was 98%.

On average, the overall accuracy of the model is close to 65%. It also says that we need better balanced data for each League index to build a more powerful predictive model.

### My best advice for future data collection-

1. The data collection should be balanced around every League Index, with the data being collected equal and big enough to build powerful predictive models
2. APM (Actions Per Minute) is a very high determinant of the League Index Ranking of a player. It is a feature that should be given ample importance for data collection.
3. More the data can be collected with diverse factors, better chances for building powerful predictive models.