Evil Geniuses

05/29/2023

Subject: Analysis of Starcraft Player Performance Dataset

Dear Stakeholders,

I am writing to present the findings of our analysis on the Starcraft player performance dataset and its potential implications for understanding and predicting player ranks. Our objective was to develop a model to predict a player's rank using the information provided in the dataset. Here is a summary of our analysis and key findings:

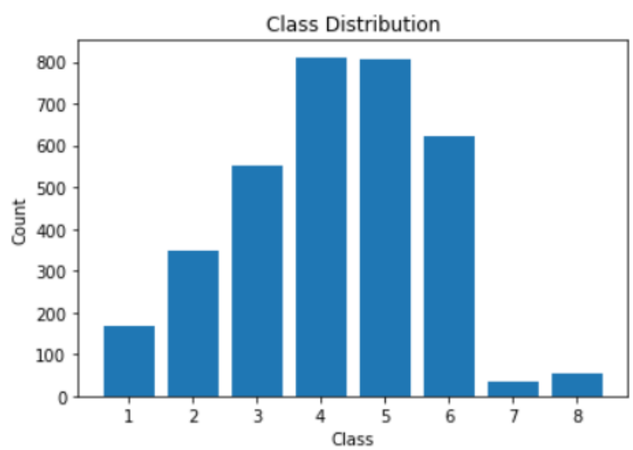
Exploratory Data Analysis:

We analyzed a dataset containing information on various aspects of player performance in ranked Starcraft games.

The dataset consisted of several features such as age, reported hours spent playing per week, actions per minute (APM), hotkey usage, minimap interactions, and more.

The target variable, LeagueIndex, represents the player's rank and is coded from 1 to 8, with higher values indicating higher ranks.

```
In [6]: class_counts = data['LeagueIndex'].value_counts()
        plt.bar(class_counts.index, class_counts.values)
        plt.xlabel('Class')
        plt.ylabel('Count')
        plt.title('Class Distribution')
        plt.show()
```



We observed an imbalanced distribution of players across different ranks, indicating that our model is biased towards specific ranks. As you can see from the image above, there are very few data with League Index of 7 and 8.

Key Insights:

```
In [19]: sorted_corr = correl['LeagueIndex'].sort_values(ascending=False)
         sorted_corr
         #found the two least correlated columns to LeagueIndex i.e. GameID and TotalHours

Out[19]: LeagueIndex             1.000000
         APM                     0.624171
         NumberOfPACs            0.589193
         AssignToHotkeys         0.487280
         SelectByHotkeys         0.428637
         UniqueHotkeys           0.322415
         WorkersMade             0.310452
         MinimapAttacks          0.270526
         TotalMapExplored        0.230347
         HoursPerWeek            0.217930
         MinimapRightClicks      0.206380
         ComplexUnitsMade        0.171190
         ComplexAbilitiesUsed    0.156033
         UniqueUnitsMade         0.151933
         ActionsInPAC            0.140303
         GameID                  0.024974
         TotalHours              0.023884
         Age                    -0.127518
         GapBetweenPACs         -0.537536
         ActionLatency          -0.659940
         Name: LeagueIndex, dtype: float64
```

Using the correlation information provided above we can see that:

1. APM 0.624171
2. NumberOfPACs 0.589193

are the two most positively correlated columns to the League Index. Similarly,

1. GapBetweenPACs -0.537536
2. ActionLatency -0.659940

are the two most negatively correlated columns to the League Index

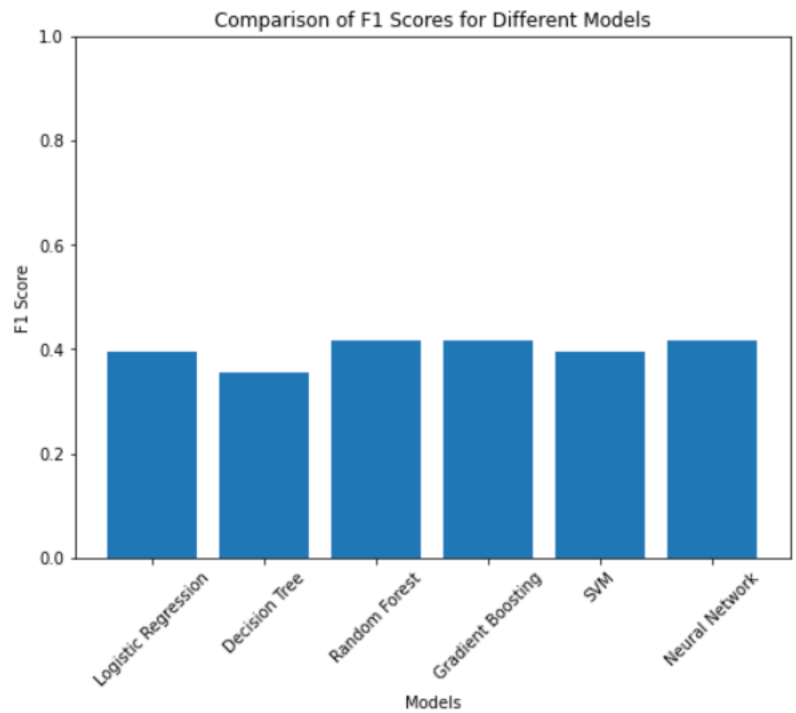
Actions per minute (APM), a measure of how quickly players execute actions in the game, exhibited a positive correlation with rank. Players with higher APM tended to have higher ranks, indicating the importance of fast decision-making and execution.

Hotkey usage, minimap interactions, and other gameplay-related metrics also demonstrated varying degrees of correlation with player ranks. These features provide valuable insights into the players' strategies and skills.

Model Building and Evaluation:

We employed advanced machine learning techniques to build a predictive model for player ranks.

After rigorous evaluation and comparison of various models, we identified a Random Forest model that demonstrates promising performance in predicting player ranks.



The selected model incorporates several features from the dataset and uses them to make accurate predictions about a player's rank. We have addressed all the issue like solving the class imbalance problem, removing the unnecessary features and handling null values.

We performed various imputation techniques, and feature engineering to make the best of the selected model.

Implications and Future Data Collection:

The results of our analysis provide valuable insights into the factors that contribute to player ranks in Starcraft.

Based on the identified correlations between certain features and player ranks, we recommend collecting additional data in specific areas to further enhance the predictive capabilities of our model.

We advise collecting more data related to gameplay strategies, player decision-making processes, and additional performance metrics that can potentially improve the accuracy of our predictions.

Based on the EDA and model results, We suggest the following:

1. Collect more samples for the minority classes: Since the dataset is imbalanced, collecting more data for the underrepresented rank levels can improve the model's performance.
2. Gather additional features: If there are relevant features that are not present in the current dataset, collecting additional data with those features can enhance the model's predictive power.
3. Monitor data quality: Ensure that the new data collection process maintains data quality standards, such as avoiding missing values, outliers, or inconsistencies.
4. Perform iterative model updates: As more data becomes available, it's beneficial to periodically update and retrain the model using the augmented dataset to capture any evolving patterns or changes in player performance.

These recommendations aim to enhance the predictive capabilities of the model and provide more accurate rank predictions.

In conclusion, our analysis of the Starcraft player performance dataset has provided valuable insights into the factors influencing player ranks. We have developed a predictive model that shows promise in accurately predicting a player's rank. With additional data collection and continued evaluation, we can further refine our model and improve its predictive capabilities.

Please feel free to reach out if you have any questions or require further clarification. We are excited about the potential of this analysis to inform decision-making and enhance our understanding of player performance in Starcraft.

Thank you for your attention and support.

Sincerely,

Nimisha Khaitan

Data Scientist

Evil Geniuses