Laldinpuia Hmar
Raghavendra Singh
CS-1390/PHY-1390-1
29 April, 2023

ASHOKA PROJECT REPORT

PROBLEM

Define a ML Problem related to Ashoka and collect (and annotate) a small dataset related to the problem

Using ML, a model can be trained to predict the likelihood that non-cis male students will join the football squad in the upcoming college season.

The management of the college, in particular the sports department, can use this information to develop strategies that will effectively motivate non-cis male students to engage in sports and advance gender variety in the college's sports programs.

DESCRIPTION

By gathering the Ashoka Premier League 5.0 and 6.0 dataset, we can predict the participation of non-cis male students in the college's football team based on a number of features, including the cost of the non-cis players (by comparing with the cost of the male players and their differences), batchwise-analysis (how many non-cis male players from each batch are there), in one tier- how many non-cis males are there compared to men proportion, etc.

We can start by data processing and cleaning to make it ready for training and then perform exploratory data analysis for better insights and provide better understanding and then we can start by training and testing the model.

ML MODEL

A target variable represented as 0 or 1, it is a binary classification with few features, so Logistic Regression model will work out the best but it can be extended to using decision trees as well. A combination of exploratory data analysis, data cleaning, preprocessing and visualization techniques, can help effectively build the classification model to predict the best outcomes. Also, as we are dealing with small datasets, we have to be cautious of overfitting and it may be necessary to use techniques like regularization or cross-validation.

DATASET-FEATURES AND PREDICTOR VARIABLES

FEATURES:
['NAME', 'Gender', 'UG', 'TIER 5.0', 'APL 5.0', 'TEAM', 'PRICE in TIER 5.0(in Million)', 'APL 6.0', 'TIER 6.0', 'PRICE In APL 6.0', 'POSITION', 'ALTERNATE POSITION', 'VISIBLE PHOTO', 'MEAN_TIER', 'COUNT_APL', 'MEAN_PRICE']

PREDICTOR VARIABLES(X):

The features are basically the players performance measure, characteristics and popularity.

1.'Gender': Categorical variable indicating whether player is male or non-cis male

2. 'UG': Continuous variable within range which will be important for batchwise analysis

3. 'MEAN_TIER': By taking mean of TIER 5 and 6, It will help identify the players ability and experience especially non-cis players

4. 'COUNT_APL': Counting whether participated in APL 5.0 or 6.0 or both which represents interest and experience in football, if it has more players with more attention, then more non-cis male will join.

5. 'TEAM': Non-cis may likely join teams which they feel more sense of belonging and connection

6. 'MEAN_PRICE': By taking mean of price in both seasons, it describes the athletic ability and will be important for analyzing cost difference with male.

7. 'POSITION'; 8. 'ALTERNATE POSITION': Effect of position preference will also account and suggests about the preferred playing positions of non-cis male players.

So, the predictor variables are:

['Gender', 'UG', 'MEAN_TIER', 'COUNT_APL', 'TEAM', 'MEAN_PRICE', 'POSITION', 'ALTERNATE POSITION']

Link to Ashoka Premier League: https://ashokapl.in/