

Starcraft 2 League Analysis

...

By Chandler Kinch

The Problem

- Starcraft 2 is a highly competitive Real Time Strategy video game made by Blizzard Entertainment
- Millions of people have played and thousands still play competitively.
- What matters most when trying to improve at Starcraft 2?
- League Placement is tricky and a key factor to good games.
- Smurfs cause poor game interactions. How can they be detected?

The Data

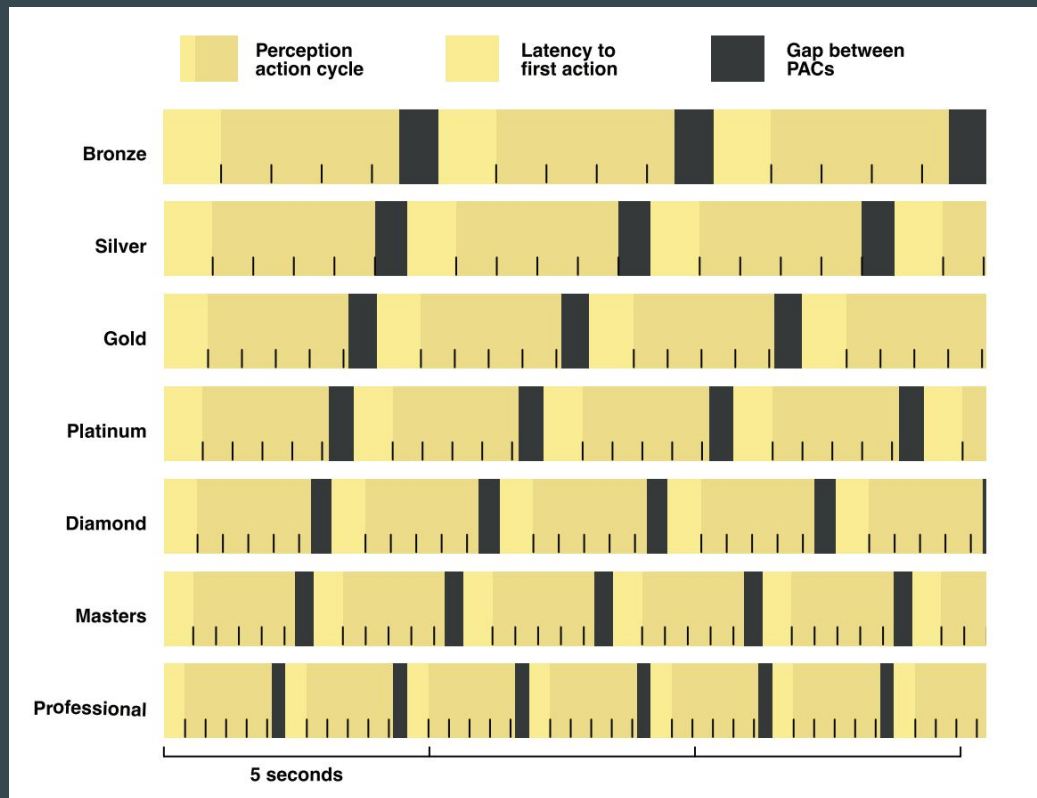
...

The Data

- Data taken from a study done at Simon Fraser University in 2013.
- This data set represents 3,395 games and includes 20 features for each game.
- Features cover gameplay aspects such as:
 - League Index(our target feature for this analysis)
 - Minimap Usage
 - Hotkey usage
 - Types of units made
 - Hours played
 - Actions per Minute(APM)
 - Data on the Perception Action Cycle

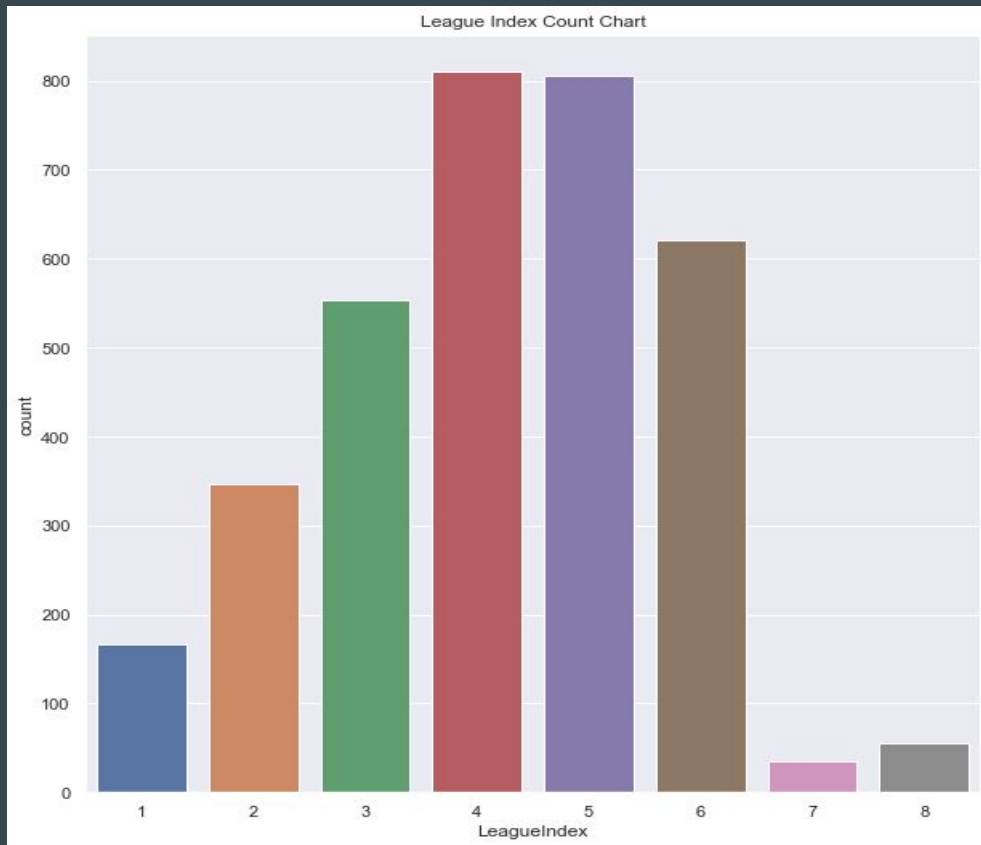
Perception Action Cycle

- Dark yellow + bright yellow = Perception Action Cycle(PAC)
- Bright yellow = Action Latency
- Black portions = Gaps between PACs
- Most aspects of PAC become faster as League increases



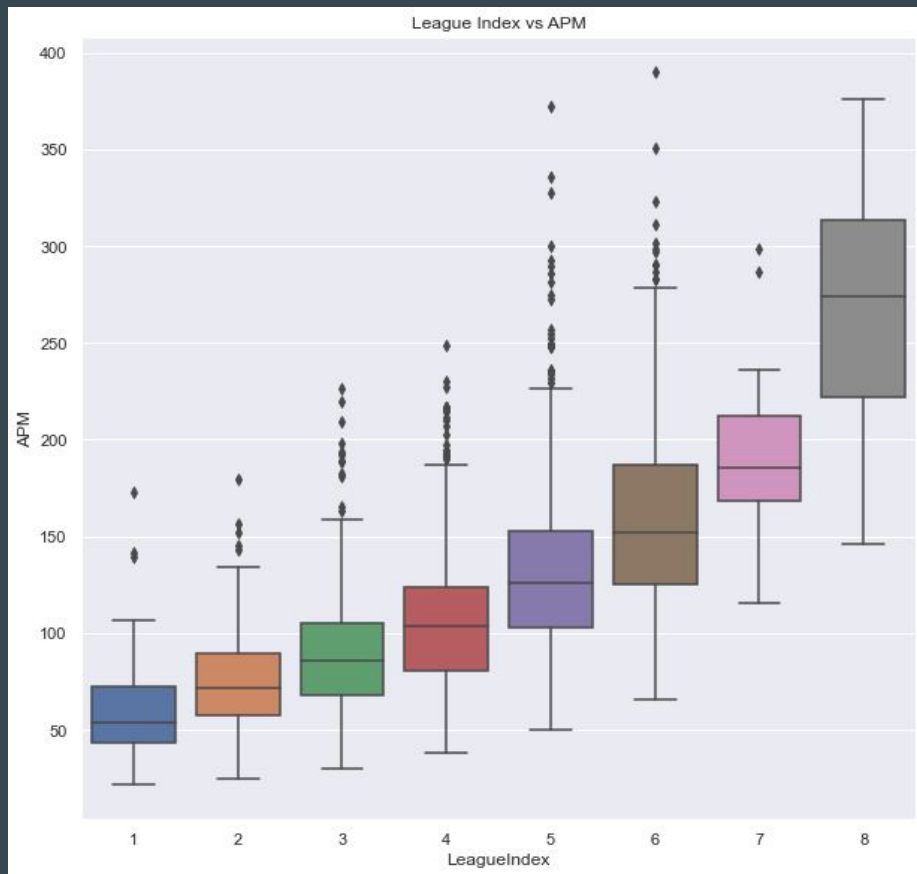
Data Spread

- Good spread of Leagues 1-6
- Lacking data in Leagues 7-8
- League 1 represents Bronze, 2 Silver and so on to 8 representing Professional games.
- Note, the highest League is Grandmaster(7). League 8 is made of Pro. gameplay.
- Most pros play in League 7.



Actions per Minute or APM

- Topic of constant discussion in the Starcraft 2 community
- High or low APM doesn't automatically determine League
- Shows important trend
- P-value of 3.31×10^{-10}

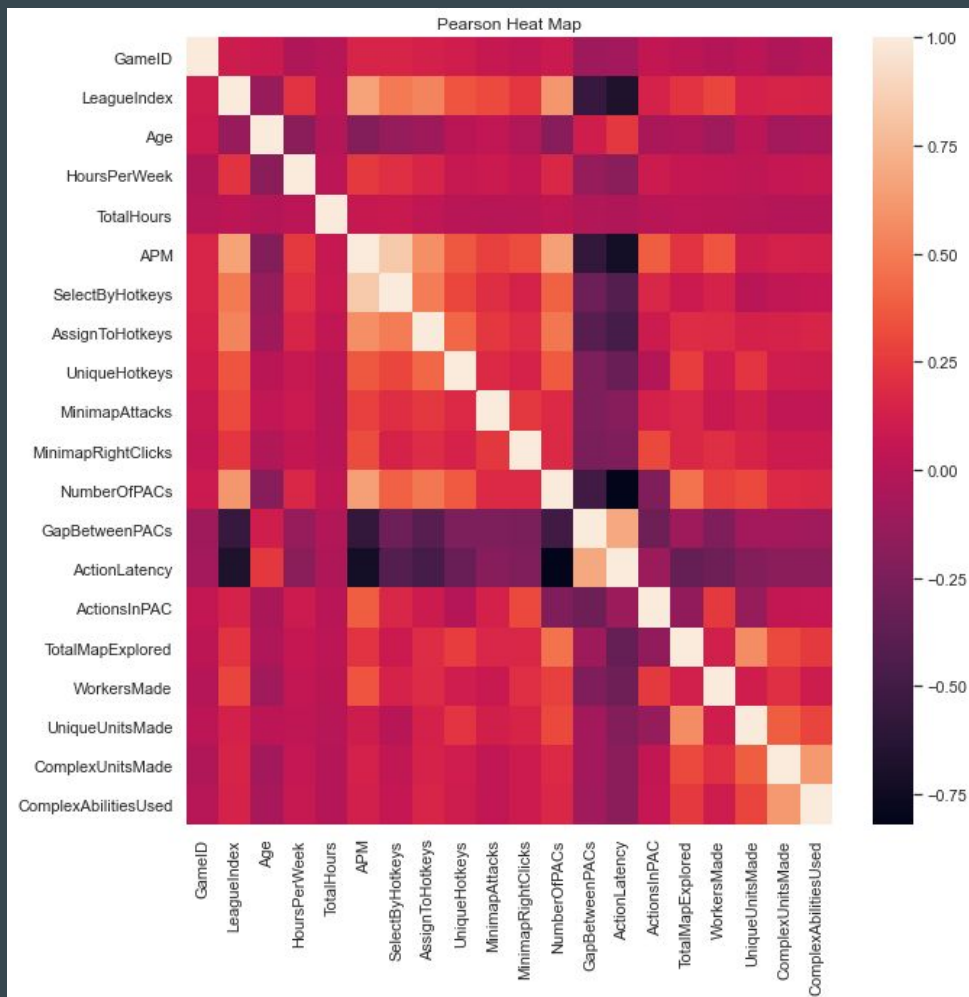


Feature Importance

...

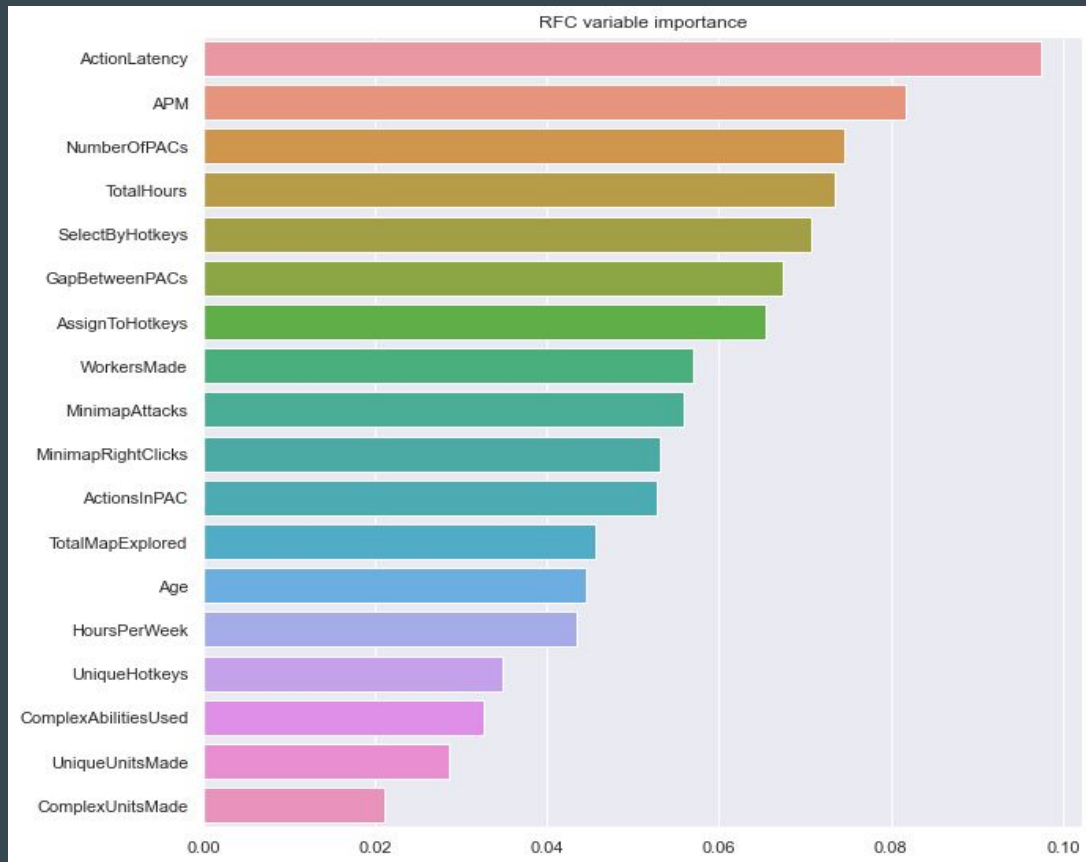
- High positive correlations with League Index:

- High negative correlations with League Index:



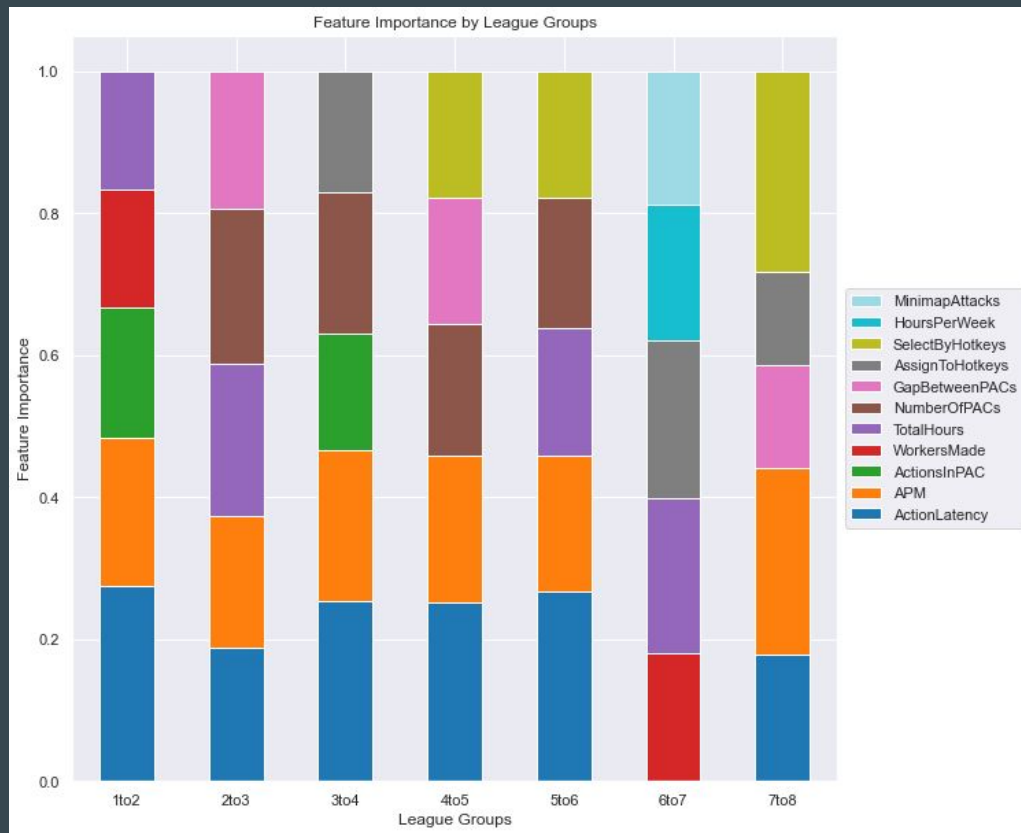
Feature Importance

- Used Random Forest Classifier
- Missing Age, Total Hours and Hours per Week for League 8, >2% of data



Feature Importance from League to League

- Action Latency and APM in all columns except 6 to 7
- Actions in PAC and Workers made appear only twice
- 6 to 7 column is very unique. May be due to lack of data.
- All variables that deal with the PAC cycle appear. Only 6 to 7 lacks any of these.



Modeling

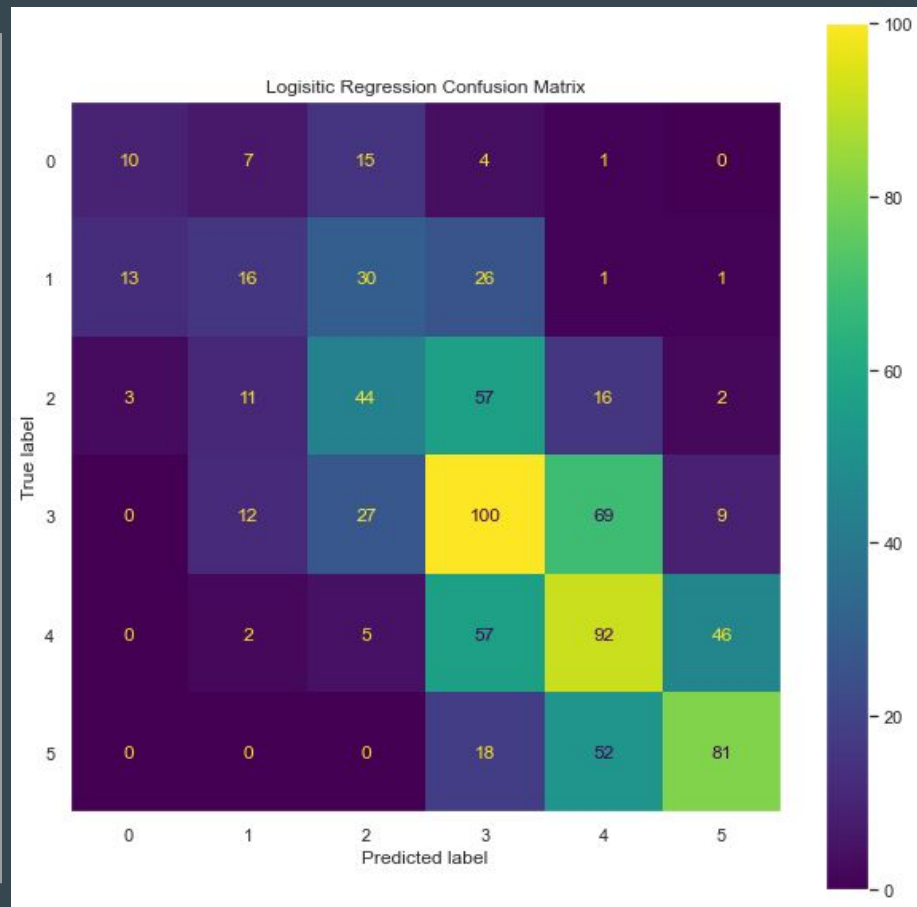
...

League Prediction Model

- Used three algorithms: K Nearest Neighbors, Logistic Regression and Random Forest
- Split data 75% train, 25% test
- Excluded Grandmaster and Professional games
- Excluded Weekly Hours played, Age and Total Hours
- Gridsearched for accuracy
- Logistic regression had the best score with 41%
- Model can predict league +/- 1 league with 86% accuracy

Model Scores

Classifier	Accuracy Score	Hyperparameter Values
K Neighbors	0.36	N_neighbors = 41
Logistic Regression	0.41	C = 0.046 Penalty = L2
Random Forest	0.39	Max_depth = 9 Max_features = auto N_estimators = 500



Smurf Detection

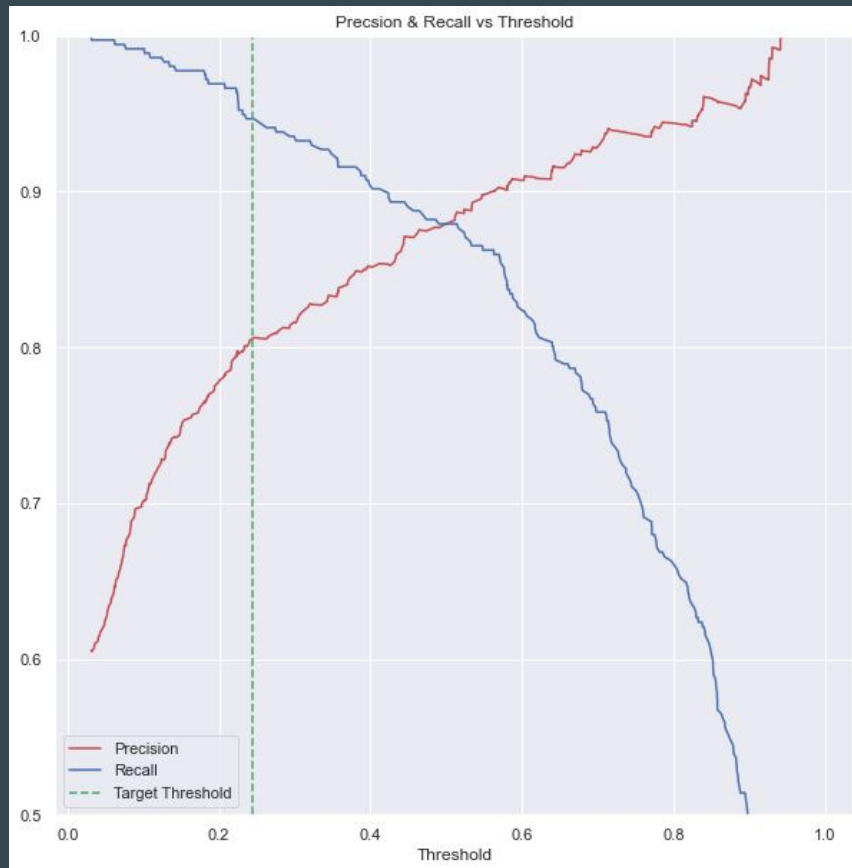
- Built three models using same three algorithms
- Set Leagues 1-3 as Low and 5-6 as High
- Exclude League 4
- Grid searched using ROC AUC
- Random Forest gave best score of 0.93

Classifier	ROC AUC Scores	Best Hyperparameters
K Neighbors	0.92	N_neighbors = 47
Logistic Regression	0.93	C = 0.359 Penalty = L2
Random Forest	0.93	Max_depth = 9 Max_features = auto N_estimators = 500

Thresholding

- Wanted to optimize recall to find as many smurfs as possible
- Threshold of 0.23 gave recall 95%

Class	Precision	Recall
Non-smurfs	0.90	0.67
Smurfs	0.79	0.95



Conclusion

- Models can be incredibly useful for ensuring fair and fun gameplay in Starcraft 2
- To answer my question of ‘What matters the most when trying to improve at Starcraft 2?’, seems speed helps. Most important features:
 - Action Latency
 - APM
 - Number of Pacs
- Features more nuanced from League to League
- Dedicated smurf data set could improve smurf detection
- Modern data could improve league placement