

StarCraft Pro Scout

Torin Rettig

Concept: Predict the Pros

Predict the next batch of StarCraft
Pro players with machine learning

What is StarCraft?

Real-Time Strategy game

Gain resources

Build an army

Destroy your opponent's
base

The original eSport



APM & PACs

APM



PACs



Motivation

eSports is a multimillion dollar industry

Tremendous growth over the past decade

Talented players are a hot commodity

ID the next star before the competition



Data

2013 SC2 Replay Analysis
study (Simon Fraser University)

3395 observations of player
game replays

19 variables: int/float

Variable Sample
Actions Per Minute
Hour Played Per Week
Complex Units Created
Perception Action Cycles
Unique Units Created
Map Explored

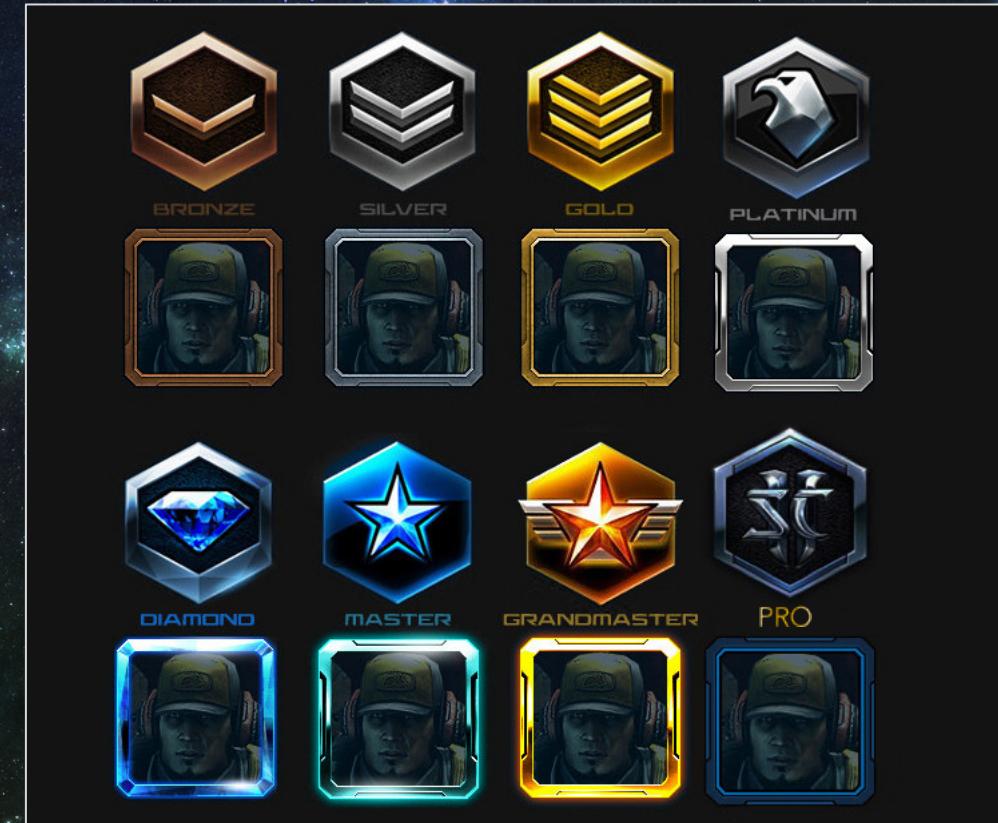
Methodology

Target: Player League Ranking

Use ML classification to predict League Ranking

Use both parametric and non-parametric modeling

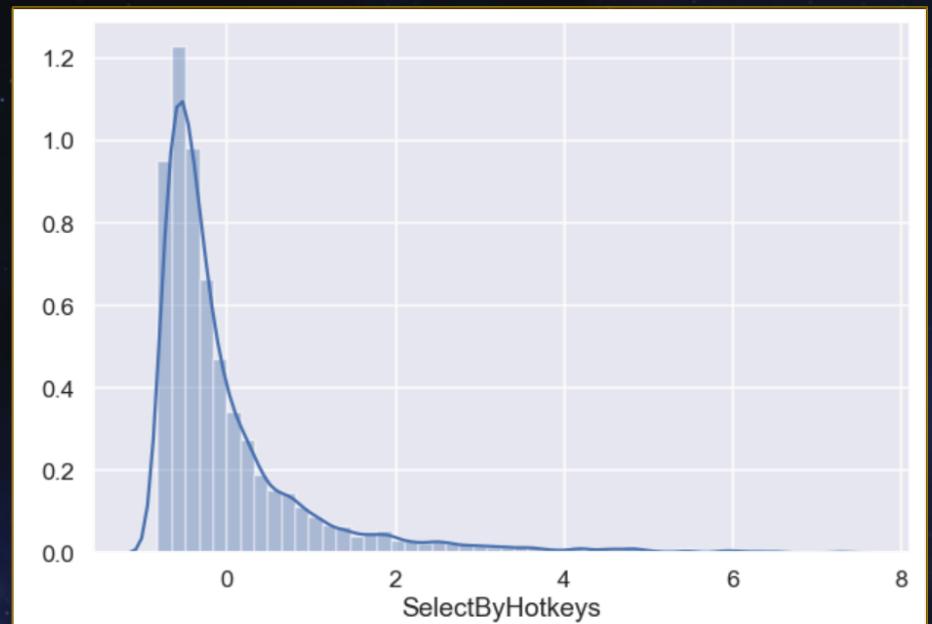
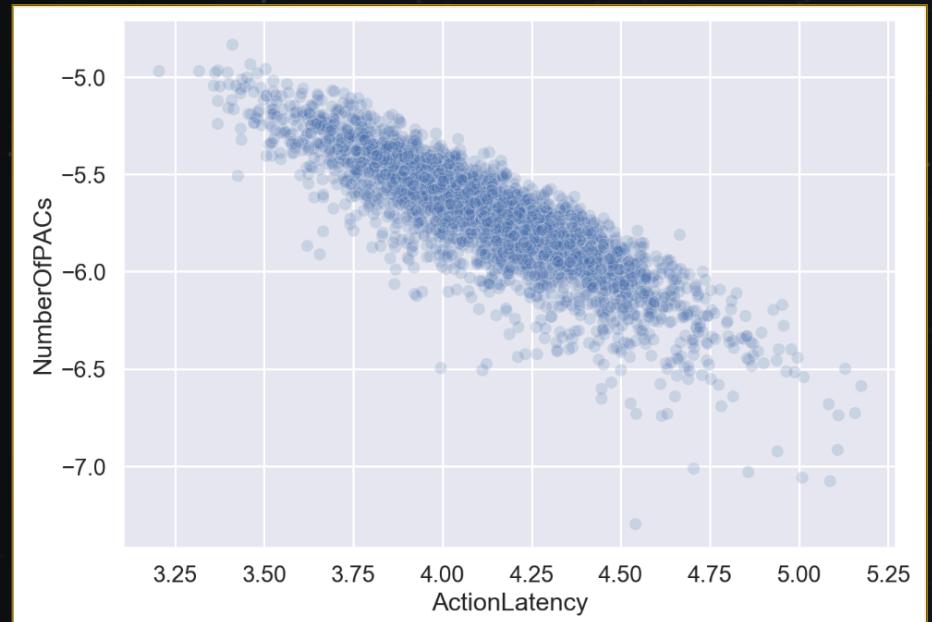
StarCraft Leagues



Feature Selection

Significant collinearity & normality issues

Created parametric & non-parametric data sets



Compress Categories

8 Leagues



Contender

2 Categories



Non-Contender



Fewer categories for increased relevance per category

Modeling

Overall goal is to maximize F1

Contender was top priority

KNN, SVC, Random Forests,
Logistic Regression

Oversampling made a big
difference

Contender

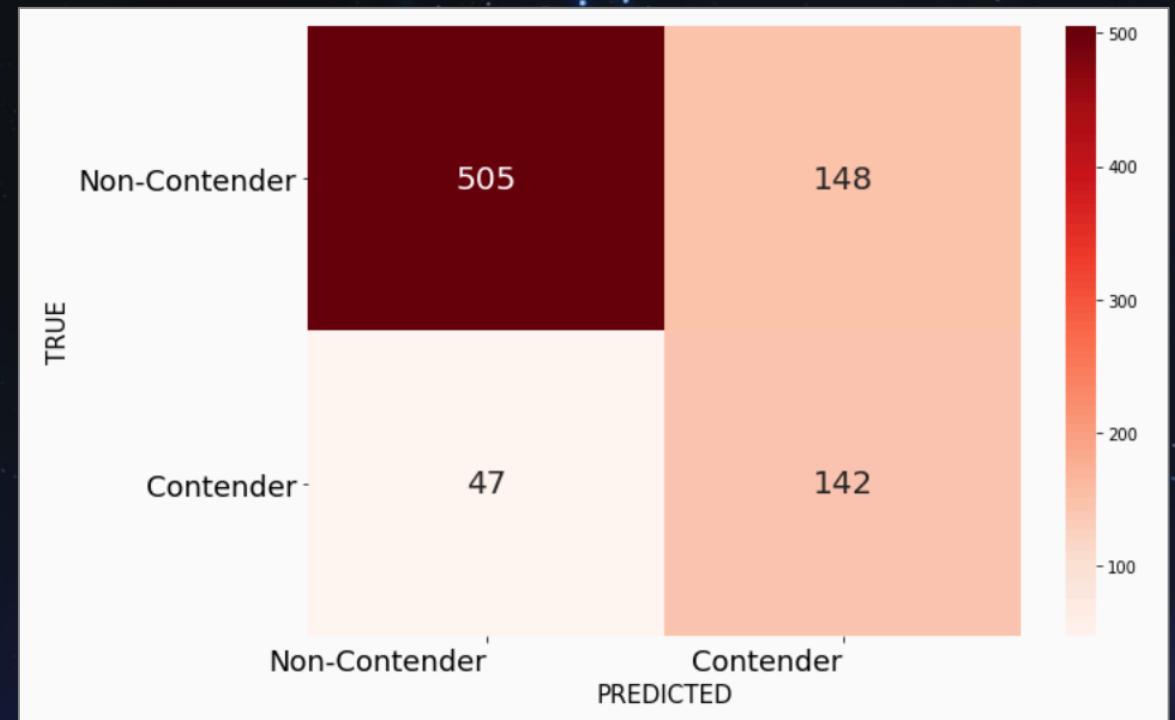


Modeling – Performance

Test Scores

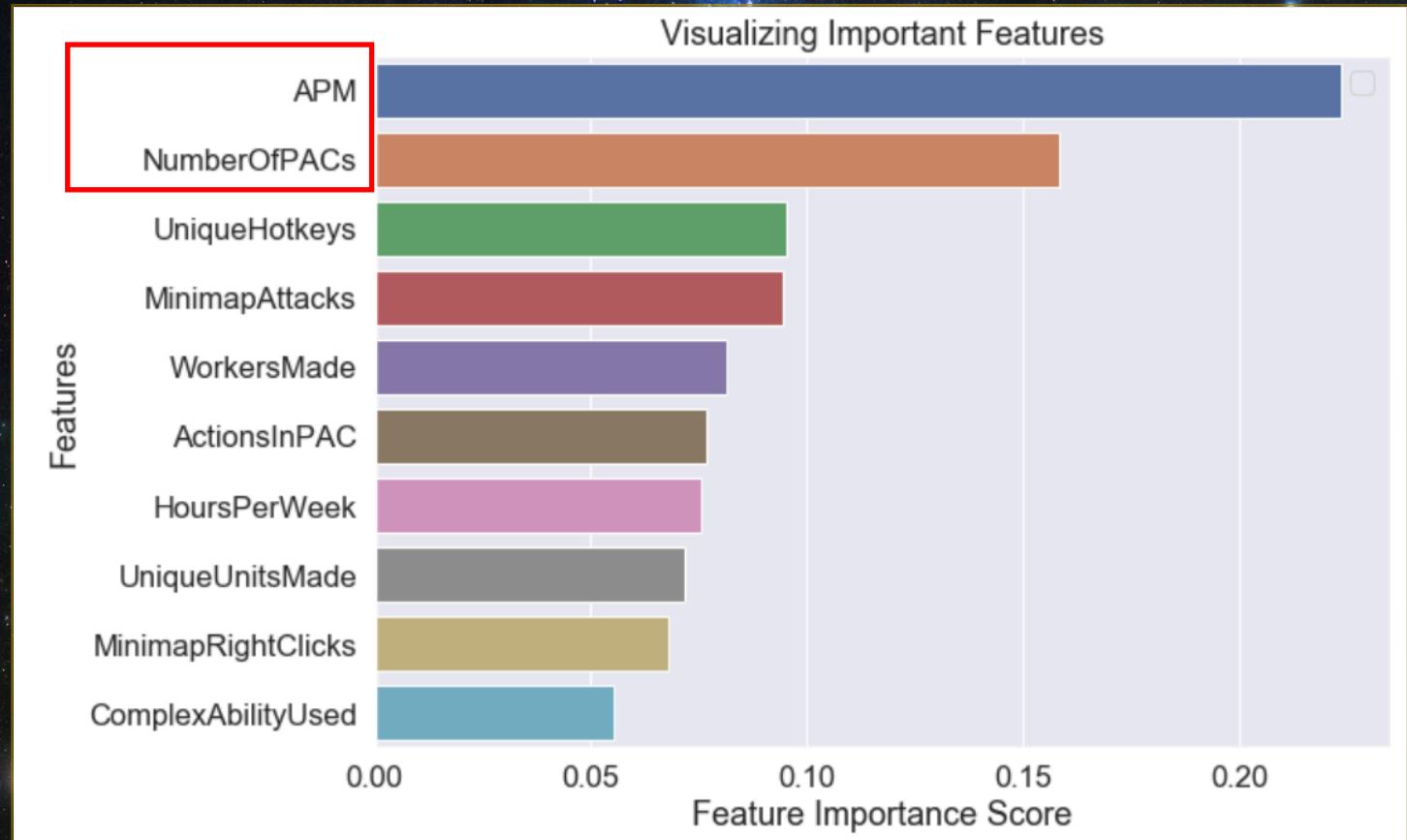
Model	F1	ROC AUC
KNN	0.5929	0.7680
SVC RBF Kernel	0.5674	0.8341
Random Forest	0.5563	0.8523
Logistic Regression	0.5913	0.8558

KNN
Confusion Matrix



Modeling – Feature Importance

Unsurprisingly both APM and PACs were the most important indicators of League placement



Conclusions

Can we predict the next SC pro player? Not reliably, yet...

There is **high variability** in player performance metrics requiring significant work to manage in modeling

Oversampling significantly improved recall performance across the board

Performance is much better than baseline, but can be much better

Future Work

Spend more time with feature engineering for parametric approaches

Try to get a larger sample size, especially for smaller categories

Apply these approaches to other competitive games



Thank You!