

Data Science and Video Games



Zach Heick

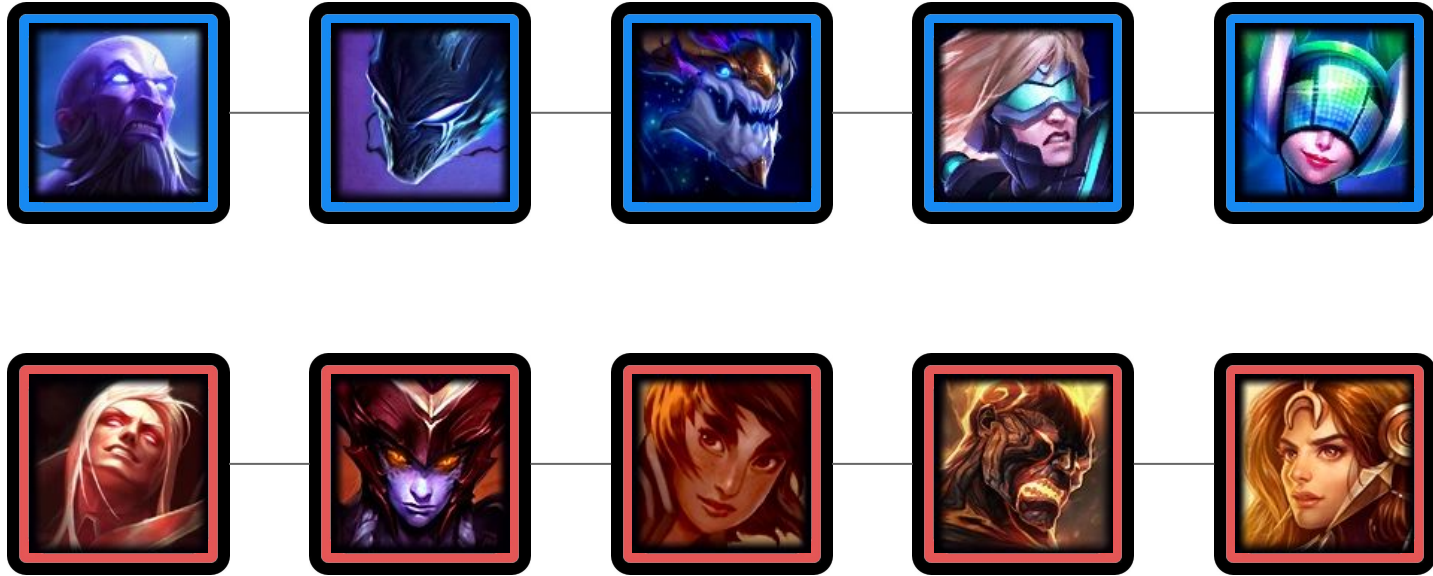
The Rise of Online Gaming

- Professional video game scene is growing rapidly
- Huge prize pools for annual and seasonal tournaments
- Teams using **data science** to provide best win conditions





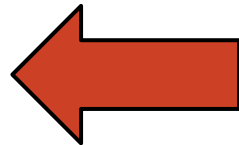
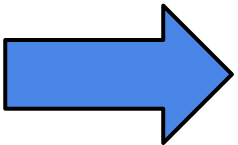
Game Rules



Two teams of 5 human players, each picks unique champion

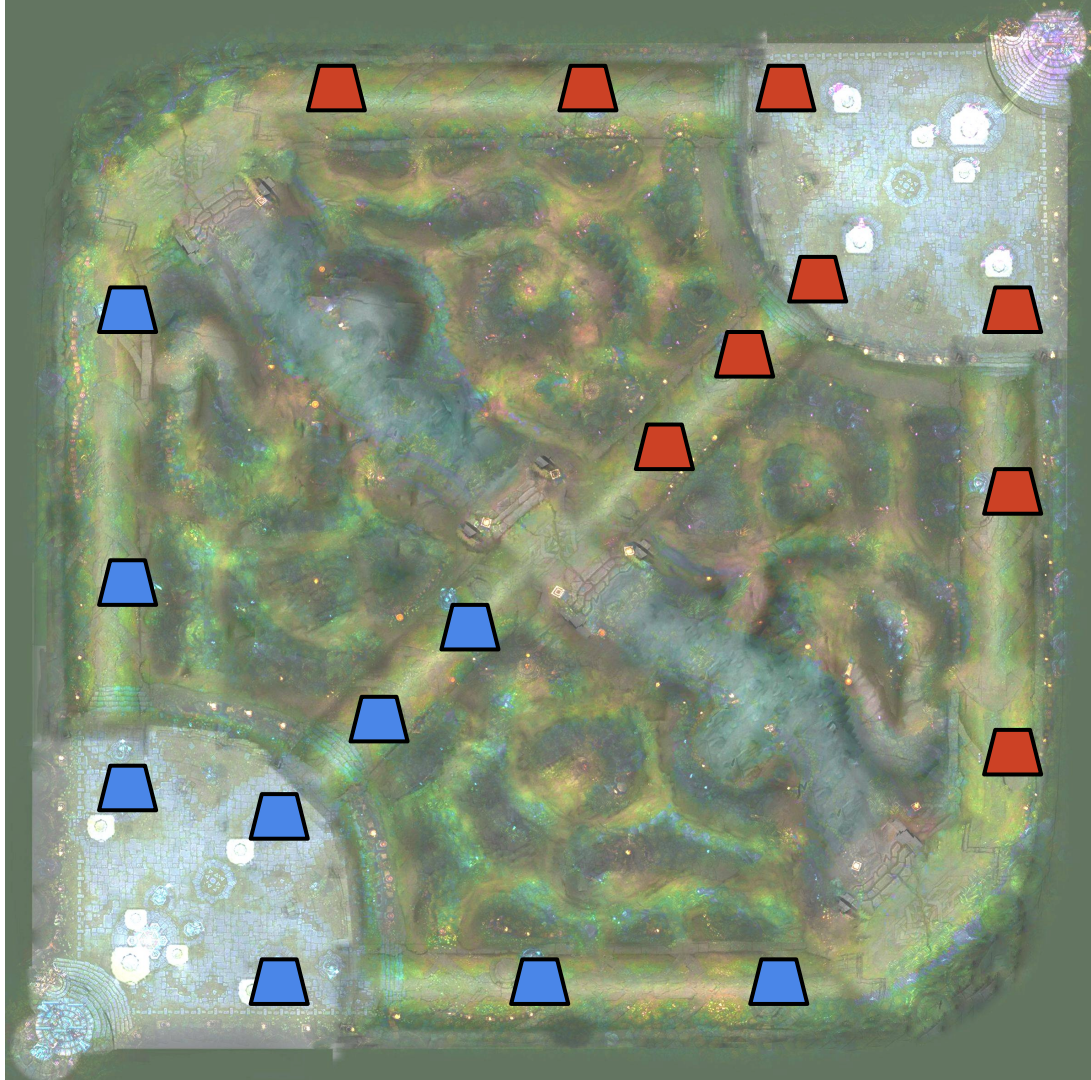


Each player also bans a champion from the other team



 Blue Turret

 Red Turret

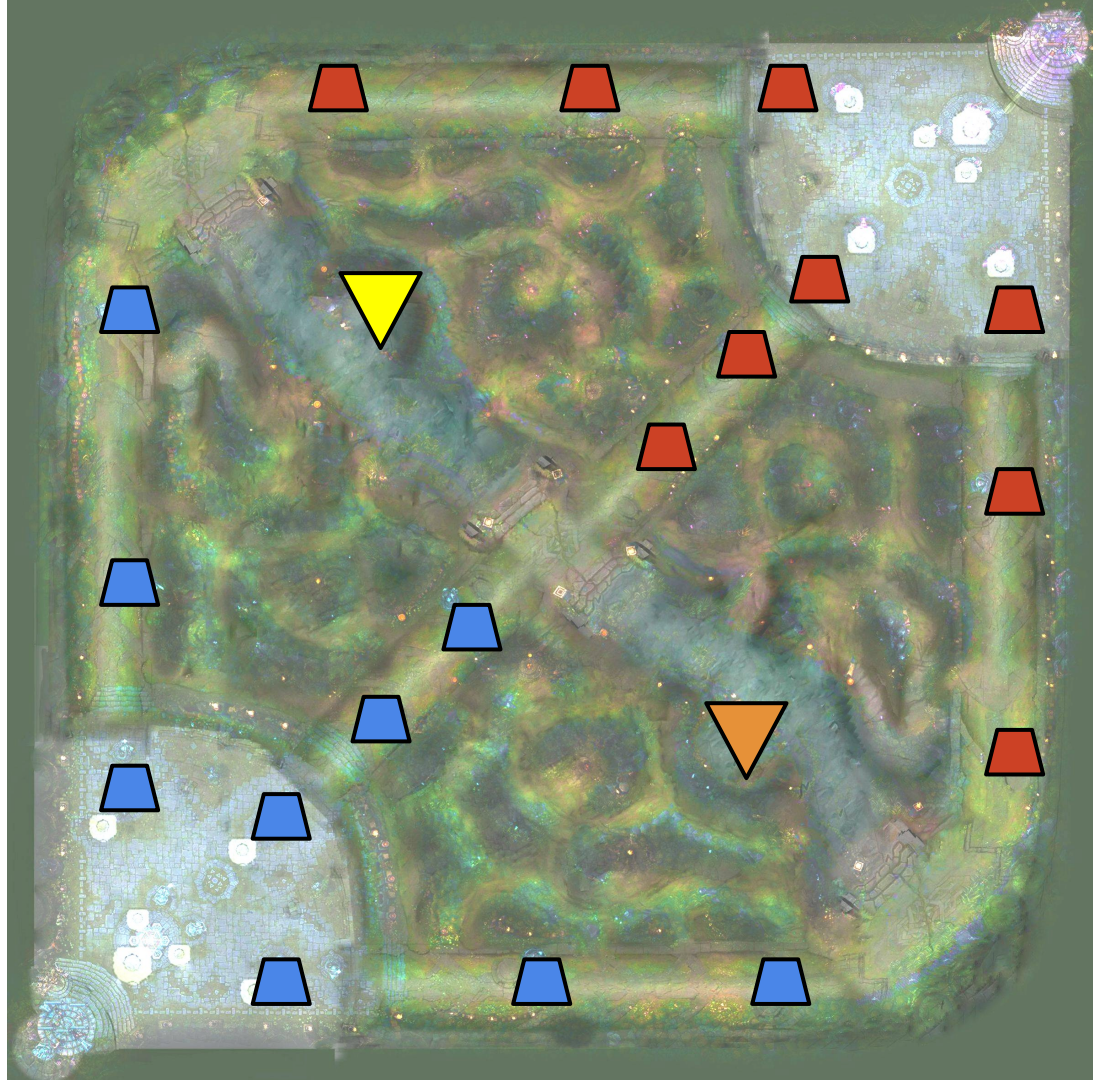


 Blue Turret

 Red Turret

 Rift

 Dragon



Features

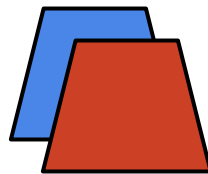


Picks

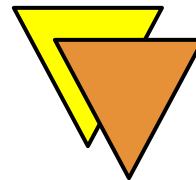
Bans



First
Kill



First
Turret



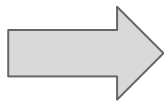
First
Rift / Dragon

= WINNER

Data and Assumptions

- Downloaded datasets from Kaggle
 - Game history containing **51490** games
 - Hero information about **138** unique heroes

kaggle™



PostgreSQL



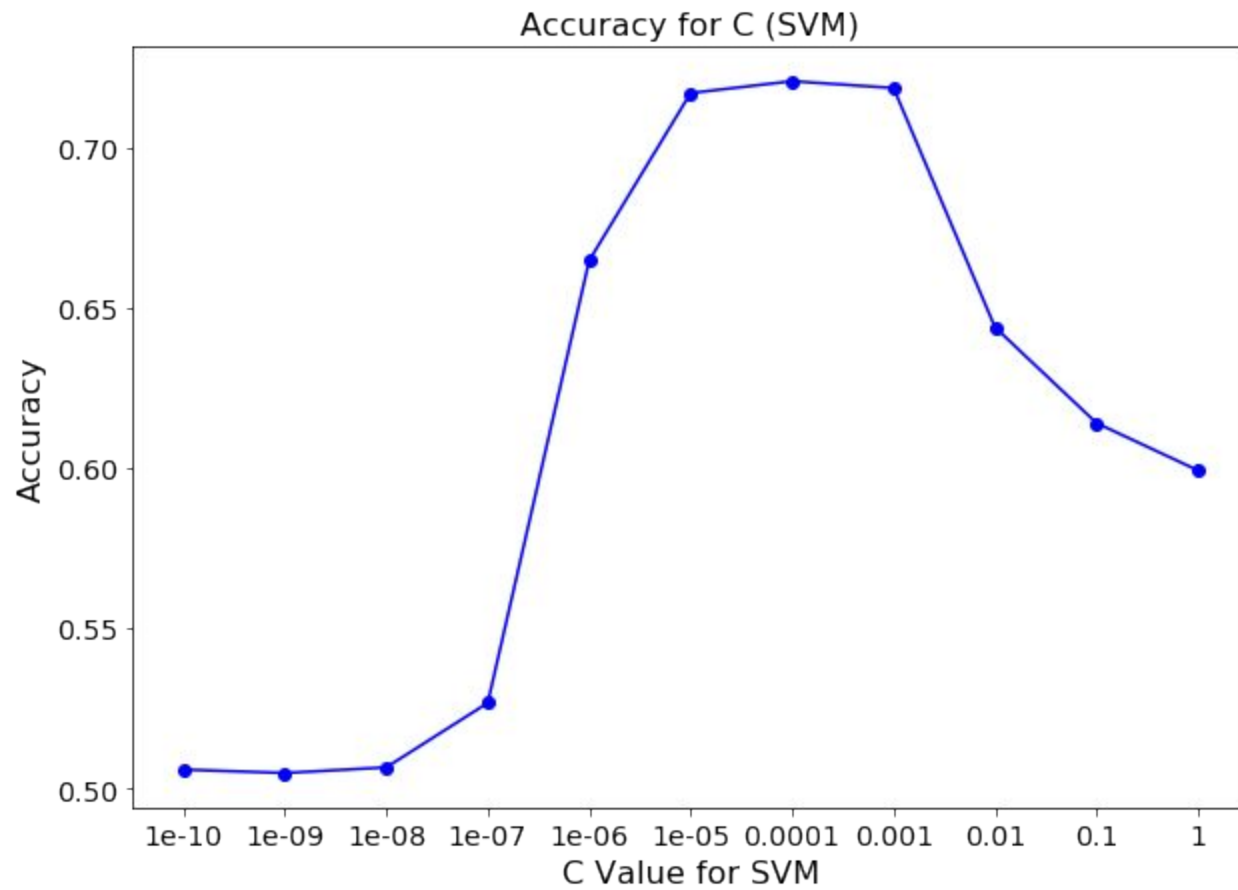
Modeling

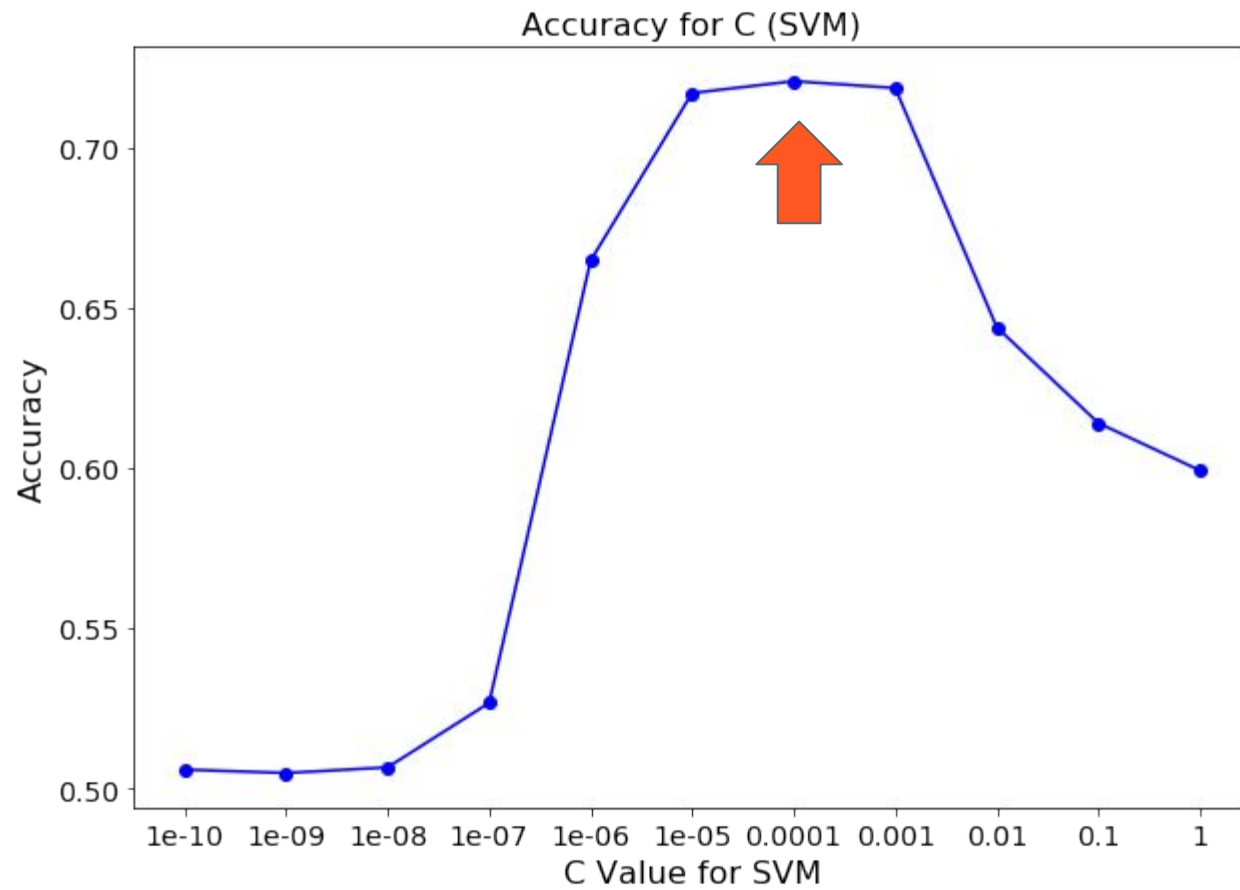
Algorithms and Hyperparameters

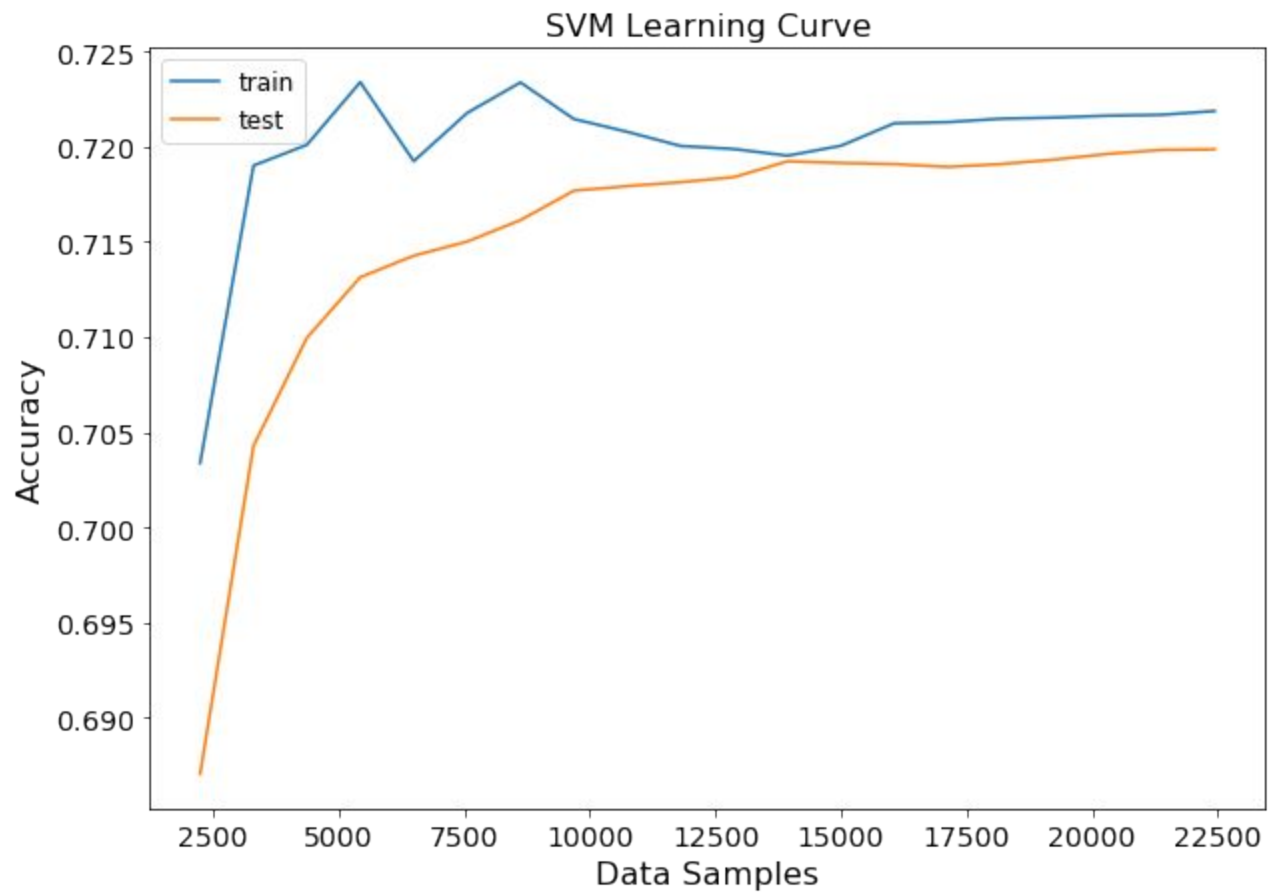
- Support Vector Machines: Budget
- Decision Tree: Depth and criterion
- Random Forest: Depth and number of trees
- Logistic Regression: Regularization Strength
- Bernoulli Naive Bayes: None

Algorithms and Hyperparameters

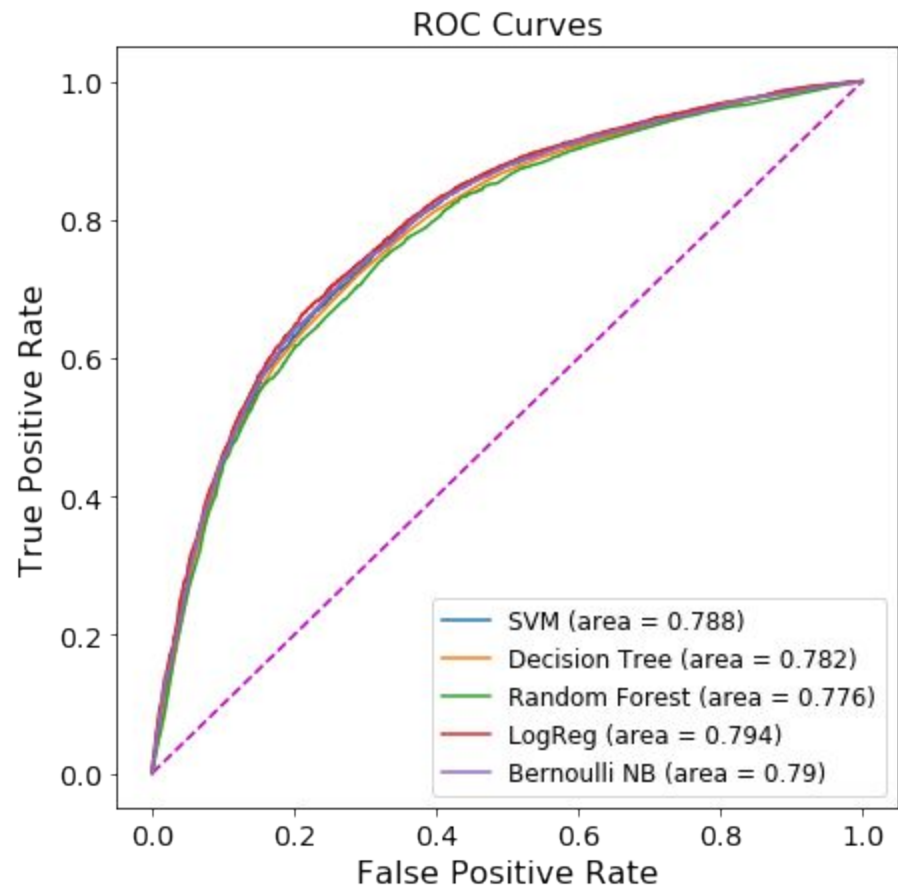
- Support Vector Machines: Budget

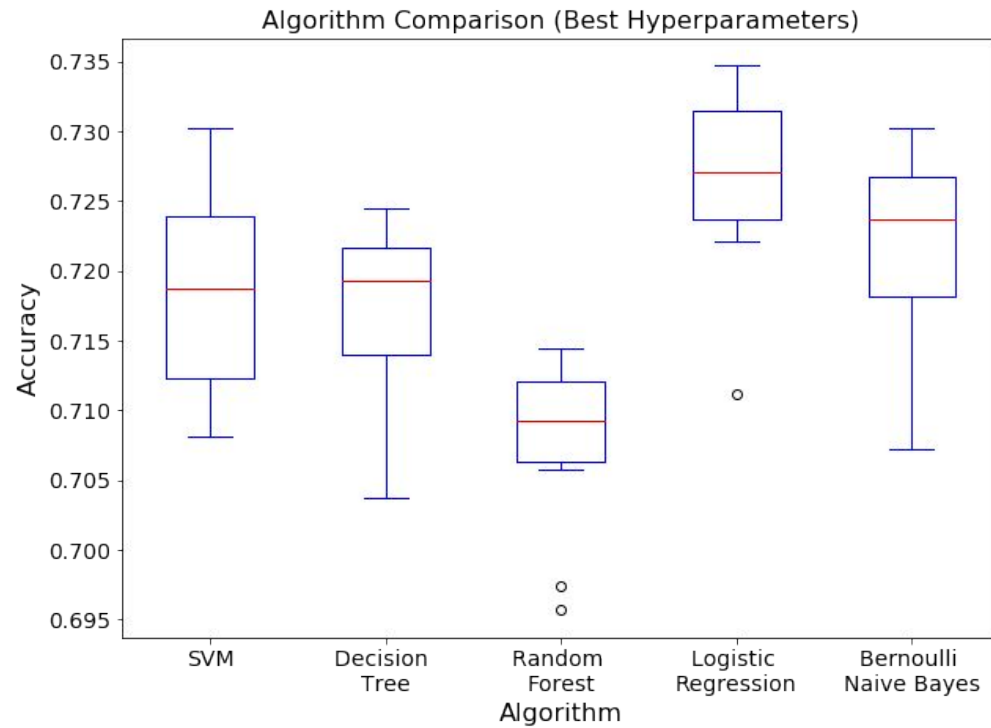











Algorithm Analysis










	Bernoulli NB	Decision Tree	Logistic Regression	Random Forest	SVM
0	10	10	1	10	10
1	14	14	10	14	14
2	15	15	14	15	15
3	16	19	15	19	16
4	19	21	19	21	19
5	21	22	21	22	21
6	22	23	22	23	22
7	23	24	23	24	23
8	24	27	24	27	24
9	27	28	27	28	27

	Bernoulli NB	Decision Tree	Logistic Regression	Random Forest	SVM
0	 10	 10	1	 10	 10
1	14	14	 10	14	14
2	15	15	14	15	15
3	16	19	15	19	16
4	19	21	19	21	19
5	21	22	21	22	21
6	22	23	22	23	22
7	23	24	23	24	23
8	24	27	24	27	24
9	27	28	27	28	27

	Bernoulli NB	Decision Tree	Logistic Regression	Random Forest	SVM
0	10	10	1	10	10
1	14	14	10	14	14
2	15	15	14	15	15
3	16	→ 19	15	→ 19	16
4	→ 19	21	→ 19	21	→ 19
5	21	22	21	22	21
6	22	23	22	23	22
7	23	24	23	24	23
8	24	27	24	27	24
9	27	28	27	28	27

	Bernoulli NB	Decision Tree	Logistic Regression	Random Forest	SVM
0	10	10	1	10	10
1	14	14	10	14	14
2	15	15	14	15	15
3	16	19	15	19	16
4	19	21	19	21	19
5	21	22	21	22	21
6	22	23	22	23	22
7	23	 24	23	 24	23
8	 24	27	 24	27	 24
9	27	28	27	28	27

How Do I Choose an Algorithm?

- Accuracy
- Interpretability
- Computation Speed
- **Support Vector Machine!**

Conclusion

- Used Support Vector Machine for final model
- 24 Features
- Final scores
 - **Accuracy:** 71.74 %
 - Precision: 71.66 %
 - Recall: 73.01 %
 - F1: 72.33 %

Thanks for listening!



LEAGUE of
LEGENDS

Appendix

The Rise of Online Gaming

- Big name sponsors and investors
- Teams using data science to provide best win conditions



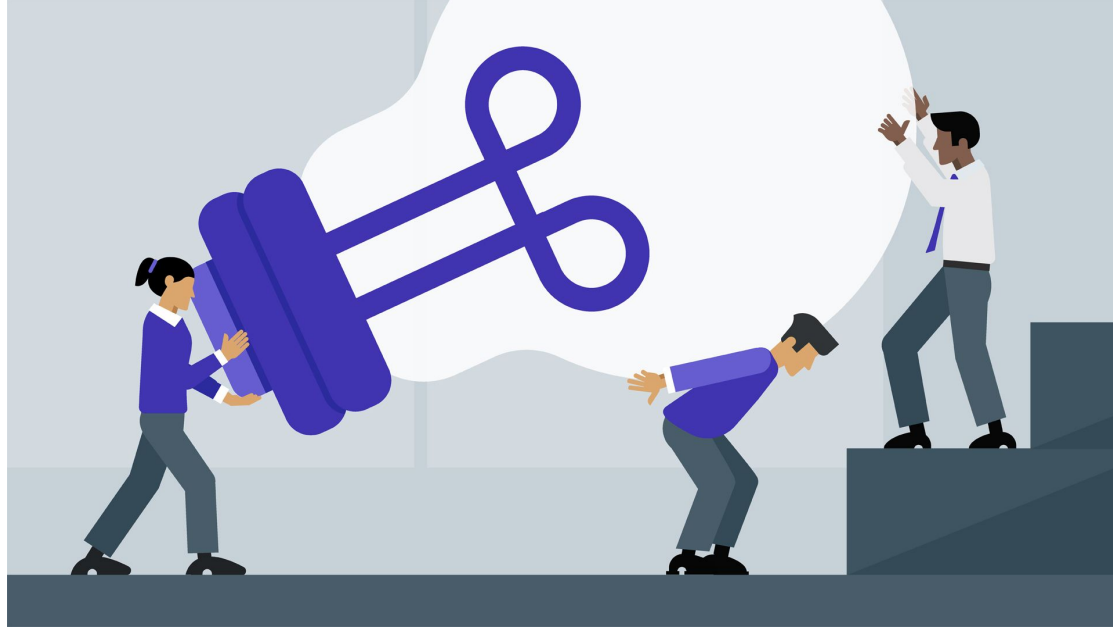
Data and Assumptions

1. The games are played on the same ranked scale.
2. Each player is playing their champion to full capacity.
3. Each player knows how to play all 138 champions at equal levels.

Modeling Pipeline

1. Use cross validation to tune hyperparameters
2. Plot learning curve
3. Create model with best hyperparameters and record scores
4. Analyze

Conclusion



Game objectives more important than team composition

Future Improvements

- Add more data regarding individual game stats
- Only look at high ranked games
- Try and measure champion synergy

