# League of Legends 2020 World Championship

Worlds 2020 Prediction Model

Scott Okamura

# League of Legends

- Online multiplayer game
  - Riot Games, 2009
- 115 million active players
- Annual $2 million+ tournament
  - Only for qualified teams

# 3,882,252

Peak online viewers during Worlds 2020

# $410,000

Average salary of LCS players 2020

# Data Science **OSEMN** Model

## Obtain

- from other location
- Query from database or API
- Extract from another file
- Generate data (eg. sensors)

## Scrub

- Filtering lines
- Extracting columns or words
- Replacing values
- Handling missing values
- Converting formats

## Explore

- Understanding data
- Deriving statistics
- Creating visualization

## Model

- Clustering
- Classification
- Regression
- Dimensionality reduction

## INterpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result

# HOW TO BECOME THE BEST TEAM IN THE WORLD

# The OSEMN Way

# Obtaining and Scrubbing Data

Oracle's Elixir

- All 2020 match data

Filtering match data

- World's 2020 participants
  - Summer split match data
- Dropping individual stats
- Imputing team stats

Creating new data table

- Average stats per team over last "x" games

**Obtain**

- from other location
- Query from database or API
- Extract from another file
- Generate data (e.g. Sensors)

**Scrub**

- Filtering lines
- Extracting columns or words
- Replacing values
- Handling missing values
- Converting formats

# Exploring and Modeling Data

| Model | Training Accuracy | Testing Accuracy | Training F1 Score | Testing F1 Score |
|---|---|---|---|---|
| LogReg | 0.769 | 0.605 | 0.798 | 0.655 |
| KNN | 0.758 | 0.630 | 0.784 | 0.681 |
| SVC | 0.785 | 0.615 | 0.819 | 0.680 |
| Tree | 1.000 | 0.505 | 1.000 | 0.548 |
| RF | 1.000 | 0.610 | 1.000 | 0.675 |
| ADA | 0.749 | 0.600 | 0.780 | 0.649 |
| GBM | 0.930 | 0.600 | 0.938 | 0.655 |
| XGB | 0.905 | 0.605 | 0.916 | 0.658 |

Baseline model scores

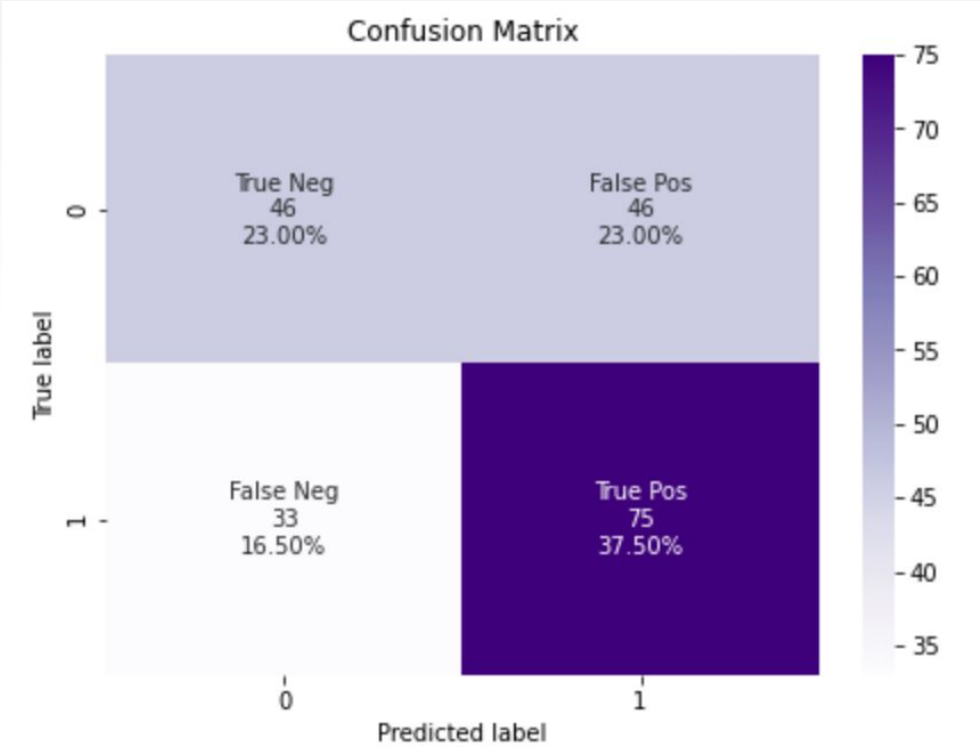
## Explore

- Understanding data
- Deriving statistics
- Creating visualization

## Model

- Clustering
- Classification
- Regression
- Dimensionality reduction

# Exploring and Modeling Data




| Model | Training Accuracy | Testing Accuracy | Training F1 Score | Testing F1 Score |
|---|---|---|---|---|
| LogReg | 0.769 | 0.605 | 0.798 | 0.655 |

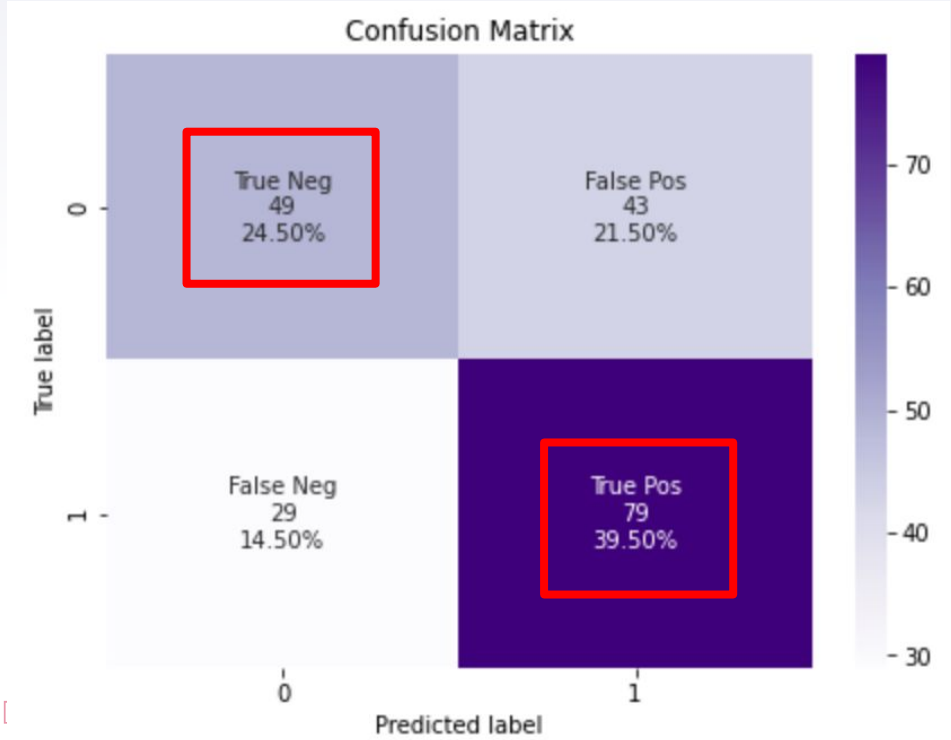
## Explore

- Understanding data
- Deriving statistics
- Creating visualization

## Model

- Clustering
- Classification
- Regression
- Dimensionality reduction

# Exploring and Modeling Data




Grid search resulted in MASSIVE improvements!


True negatives:

- 46 → 49

True positives:

- 75 → 79

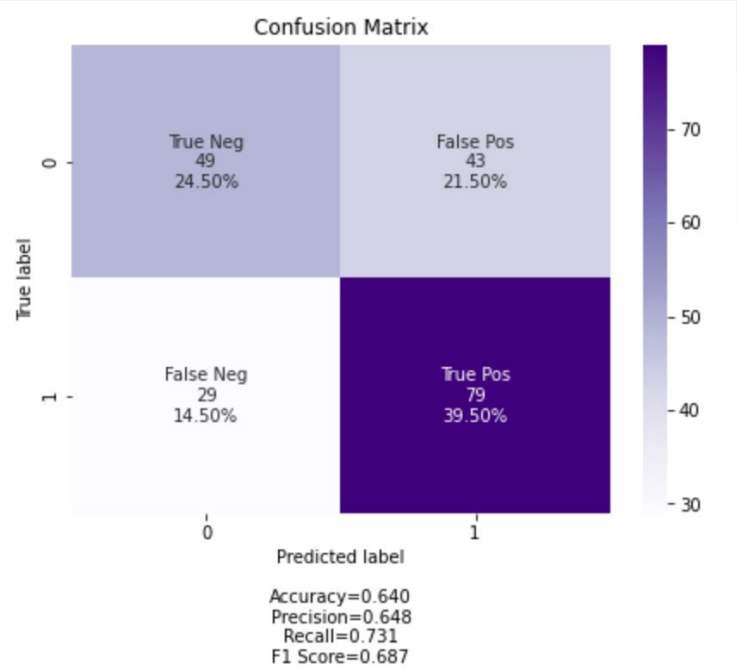

- KNeighbors Classifier: k=5, leaf_size=20

# Exploring Data

| Model | Training Accuracy | Testing Accuracy | Training F1 Score | Testing F1 Score |
|---|---|---|---|---|
| Logistic Regression | 0.798000 | 0.610000 | 0.819000 | 0.655000 |
| K Nearest Neighbors | 0.760000 | 0.640000 | 0.787000 | 0.687000 |
| Support Vector Machines | 0.788000 | 0.610000 | 0.815000 | 0.669000 |
| Decision Tree | 0.706000 | 0.575000 | 0.721000 | 0.608000 |
| Random Forest | 0.699000 | 0.605000 | 0.768000 | 0.706000 |
| AdaBoost | 0.721000 | 0.610000 | 0.762000 | 0.678000 |
| Gradient Boost | 0.918000 | 0.545000 | 0.928000 | 0.613000 |
| XGBoost | 0.979000 | 0.560000 | 0.981000 | 0.627000 |

GridsearchCV scores

# Interpreting Results




K Neighbors Classifier

64% accuracy

Decent but needs work

**IN**terpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result

# Interpreting Results

Recommendation:

Model needs to perform better to be used by coaches.

However, can be used recreationally by fans of the game.

**IN**terpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result

# Future Work on Models

Increase the number of total data points

Optimize functions to improve results

Adjust hyperparameters to possibly improve model performance

# Future Work for Team

Build a model that can predict the winner of a match with a high degree of accuracy

Find which stats (features in model) the elite teams outclass others in

Raise quality of practice by strategizing around key stats

# Thank You

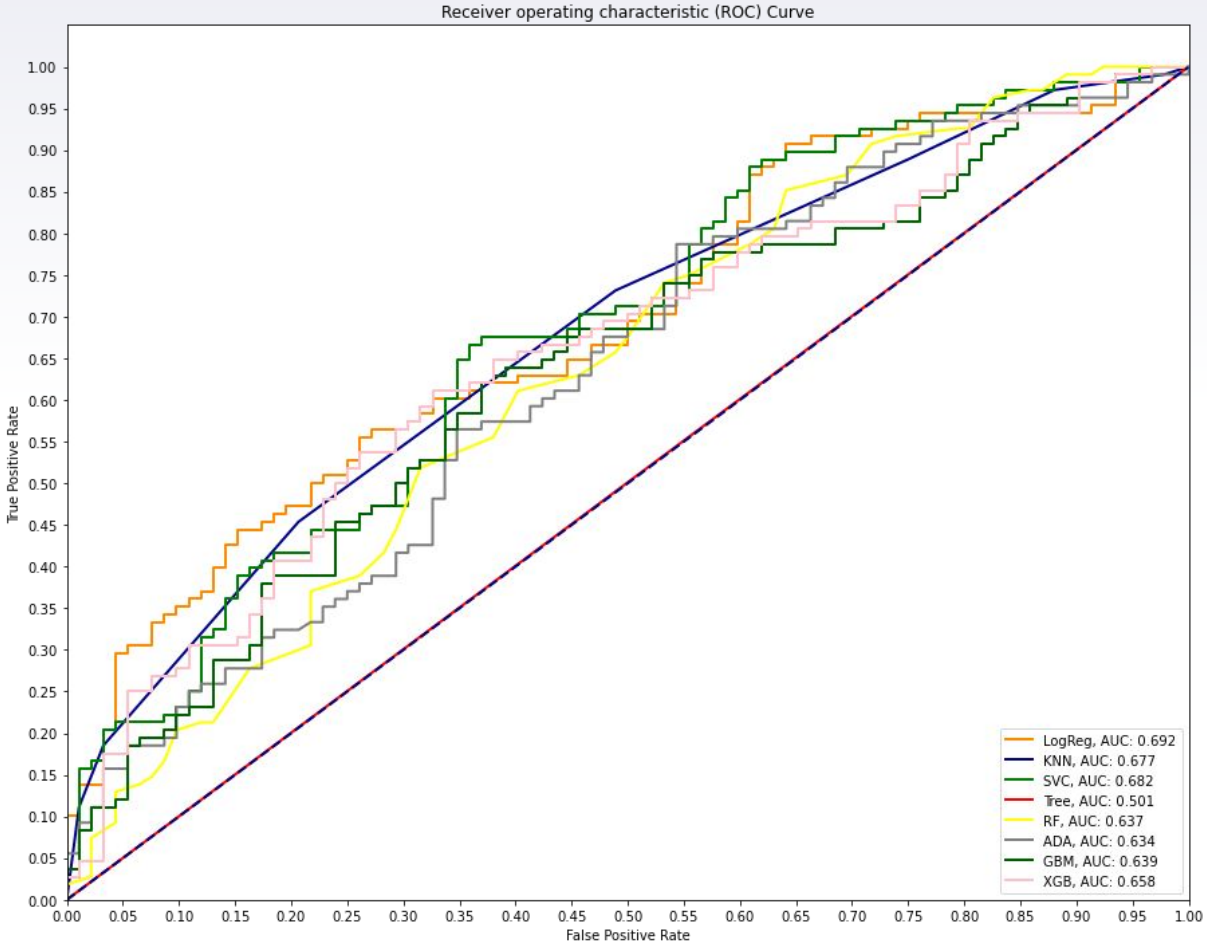
scott.okamura@gmail.com

Slidesgo

Flaticon

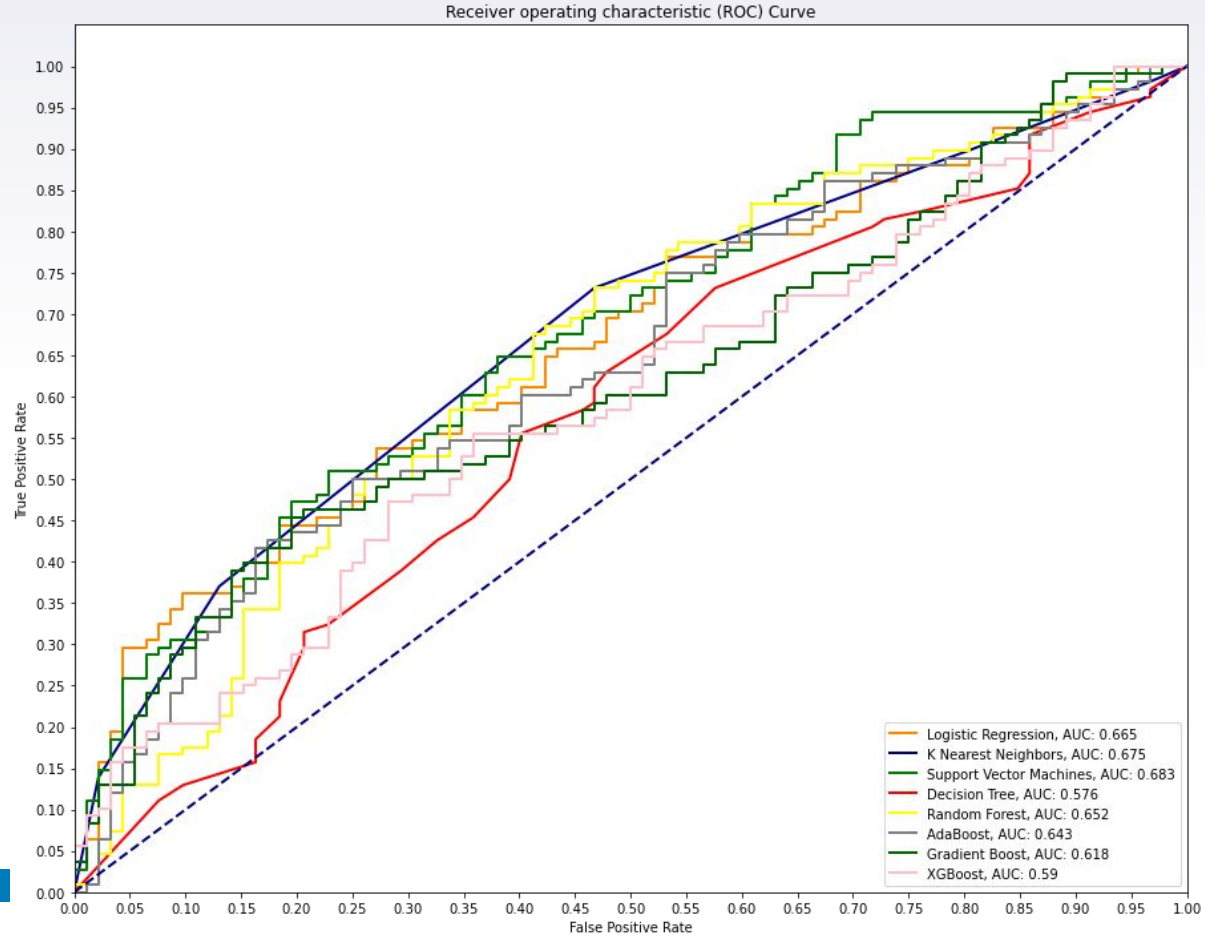
Freepik

# Appendix 1




Receiver operating curves for baseline models

# Appendix 2




Receiver operating curves for improved models