

# Winner Winner Chickie Dinner?

DSI26 Capstone: Winner Prediction for Fortnite

Presented by Yong Fah Aik

x x x x x  
x x x x x  
x x x x x  
x x x x x  
x x x x x



# Table of contents

**01**

**About the project**

**02**

**Data & Workflow**

**03**

**Data Cleaning & EDA**

**04**

**Models**

**05**

**Evaluation**

**06**

**Conclusion**

x x x x x  
x x x x x  
x x x x x  
x x x x x  
x x x x x

# 01

## About the project



# Overview



Electronic sports (or esports) is a form of competition utilizing video game. Esports differs from regular video gaming in that it is competitive (human-vs-human) and, like traditional sports, usually has an interesting spectator element.

# Genres of Esports

- Multiplayer Online Battle Arena (MOBA)
- First-Person Shooter (FPS)
- Battle Royale
- Real-Time Strategy (RTS)
- Fighting
- Collectible Card Games (CCG)
- Sports & Racing



# 1 billion



Global Revenue for  
Esports in 2020

# 495 million



Global Audience for  
Esports in 2020



# 125 million



Prize money awarded in  
tournaments in 2020



"Our world-class events and digital infrastructure have also made Singapore an attractive location for the industry to hold gaming and e-sports events here. The government will continue to support companies as they push boundaries through experimenting with new and immersive content formats and business models, as well as level up the quality of our local talent to become leading creators of world-class content."

**—Singapore Tourism  
Board and Enterprise  
Singapore**



# Singapore Focus



14 – 17 October 2021

A Hybrid Event

Held in Singapore and Online





# GEFestival



**Global  
Esports**  
FEDERATION



**GLOBAL  
ESPORTS  
GAMES  
2021  
SINGAPORE**



# Top 3 Games Awarding Prize Money

#1

*Dota 2*

\$280,180,096.65

4209 Players

1605 Tournaments

#2

*Counter-Strike: Global Offensive*

\$131,020,704.97

14734 Players

6103 Tournaments

#3

*Fortnite*

\$111,339,316.05

4931 Players

750 Tournaments

According to  
Esports Earnings.

# Problem Statement

- Predicting the winner of a match using historical player data



## Genre

Battle Royale – Fortnite



## Scope

Solo Matches, where the winner competes with up to 99 other players



## Model

Classification-based model to predict winner





# FORTNITE









# 02

## Data & Workflow

# Workflow

**Web Scraping**



**Data Cleaning  
& EDA**



**Classification  
Modelling**



**Evaluation &  
Conclusion**



xx xx xx xx  
xx xx xx xx  
xx xx xx xx  
xx xx xx xx



# Web Scraping

First set of data to be gathered from  
Fortnite Tracker, Match Statistics, using:

- Requests library
- Selenium Webdriver







# Web Scraping

Second set of data to be gathered from  
Fortnite Tracker, Player Statistics, using:

- Fortnite Tracker API

x x x x x  
x x x x x  
x x x x x  
x x x x x

# Data: Datasets



## Match Statistics

Critical information:

- Eliminations
- Points
- Time Alive
- Placement

Misc info such as:

- Team ID
- Event ID
- Geo Identities, etc.



## Player Statistics

Cumulative Statistics:

- Matches
- Kills
- Wins
- Score
- Minutes Played
- Top Rankings, etc.

Ratio Statistics

- Kills per game / per minute
- Win ratio
- Score per match / per minute
- Average Time Played
- Kills to deaths ratio



# Data: Player Statistics - Match Modes



## Solo

Solo Player



## Trios

A Maximum of  
a Team of 3



## Lifetime

Cumulative Statistics  
of All Modes



## Duos

A Maximum of  
a Team of 2



## Squads

A Maximum of  
a Team of 4



## LTM

Limited Time Modes  
with Special Rules, not  
counted in Lifetime



LAST 7 DAYS

LAST 30 DAYS

LIFETIME

56,500 Matches

 BEST TRN RATING  
**4,999** **LEGEND**

SCORE  
**23,250,105**

TOP 3/5/10  
**40,115**

TOP 6/12/25  
**45,533**

TIME PLAYED   
**454D 6H 54M**

# Sample Player Statistics

WINS  
**33,135**



WIN %  
**58.60**

KILLS  
**359,991**

K/D  
**15.41**

## SOLO

21,539 Matches

 TRN RATING   
**4,987** **LEGEND**

RANK  
-

WINS  
**9,362**

KILLS  
**138,680**

WIN %  
**43.50**

K/D  
**11.39**

TOP 10  
**13,163**

TOP 25  
**16,242**

TIME PLAYED  
**97D 15H 42M**

AVG. MATCH TIME  
**6M 31S**

KILLS/MATCH  
**6.44**

KILLS/MIN  
**0.99**



SCORE/MATCH  
**225.87**

SCORE/MIN  
**34.60**

SCORE  
**4,864,983**

## DUOS

4,824 Matches

 TRN RATING   
**4,999** **LEGEND**

RANK  
-

WINS  
**2,132**

KILLS  
**26,023**

WIN %  
**44.20**

K/D  
**9.67**

TOP 5  
**2,813**

TOP 12  
**3,355**

TIME PLAYED  
**37D 23H 50M**

AVG. MATCH TIME  
**11M 20S**

KILLS/MATCH  
**5.39**

KILLS/MIN  
**0.48**



SCORE/MATCH  
**359.96**

SCORE/MIN  
**31.74**

SCORE  
**1,736,445**

## SQUADS

23,281 Matches

 TRN RATING   
**4,983** **LEGEND**

RANK  
-

WINS  
**16,248**

KILLS  
**145,658**

WIN %  
**69.80**

K/D  
**20.71**

TOP 3  
**18,342**

TOP 6  
**19,810**

TIME PLAYED  
**231D 15H 1M**

AVG. MATCH TIME  
**14M 19S**

KILLS/MATCH  
**6.26**

KILLS/MIN  
**0.44**

SCORE/MATCH  
**529.16**

SCORE/MIN  
**36.94**

SCORE  
**12,319,410**

The background is a solid dark purple. It features several abstract elements: a large, light purple, rounded shape on the left; a jagged, light purple shape in the top right; a dark purple jagged shape in the bottom left; a series of four small orange triangles pointing up and to the right on the left side; a horizontal line with three small white circles at its ends and one in the middle; a 4x4 grid of small white 'x' marks in the bottom center; and a dark purple, rounded rectangular shape on the right with a white oval and a small white triangle. Scattered throughout are several small white circles.

# 03

## Data Cleaning & EDA

# Data Cleaning

x x x x x  
x x x x x  
x x x x x  
x x x x x  
x x x x x

## Match Statistics

- Missing data: Dropping said rows as the data is faulty
- Dropping Miscellaneous Columns such as IDs, Insert Time and End Time that is non-float

## Player Statistics

- Missing data: Dropping Ratings columns and dropping rows with mostly empty values
- Dropping Zero value columns

## Merger of Match Statistics and Player Statistics

# Feature Engineering

x x x x x  
x x x x x  
x x x x x  
x x x x x



## Player Statistics

### Lifetime Statistics

- Feature columns available in the individual sections but not provided (Mostly the ratio statistics)
- Making the Top statistics more interpretable (Top 3/5/10 & Top 6/12/25)



1.44%

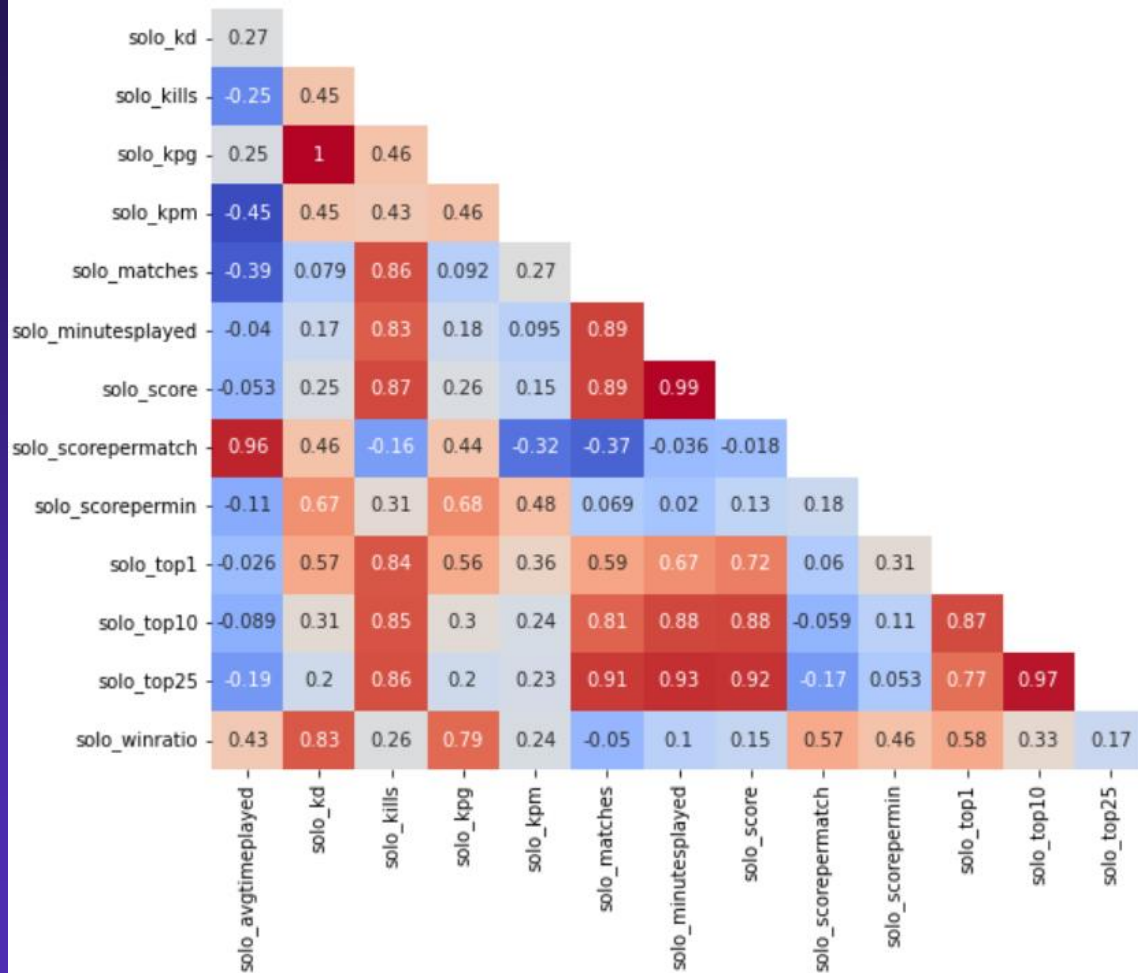
Winner Count in Data



# Correlation

## Solo Statistics

- Unsurprisingly, all of the columns are heavily correlated with one another
- The correlation for match modes are also similar.



# Correlation

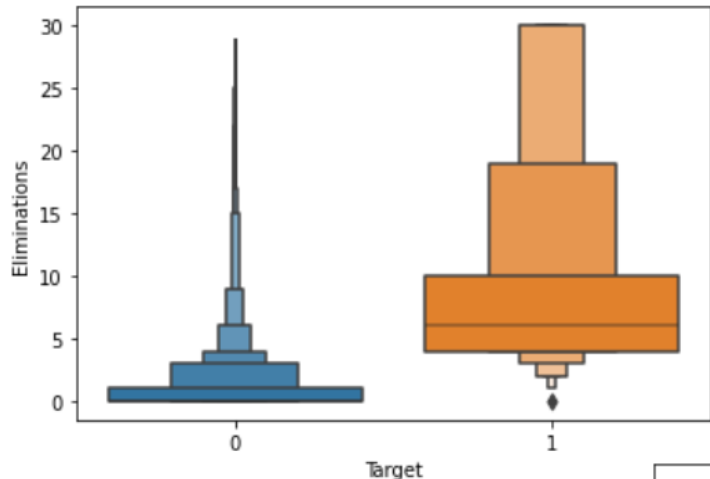
## Correlation to the Winner

- The 3 match statistics **'eliminations'**, **'points'** and **'timealive'** are more correlated compared to the player statistics.
- Very low correlations for the player statistics  
=> Model not being very effective

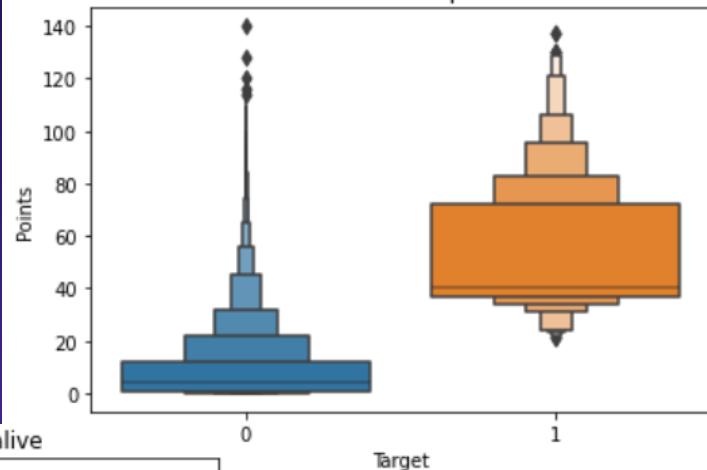
target	1
eliminations	0.36
points	0.36
timealive	0.11
trios_kpg	0.065
trios_kd	0.063
trios_scorepermatch	0.059
trios_winratio	0.057
lifetime_kpg	0.057
lifetime_kd	0.056
solo_kd	0.056
solo_kpg	0.056
trios_avgtimeplayed	0.053
duos_kpg	0.053
duos_kd	0.052
solo_winratio	0.051
trios_kpm	0.048
trios_top1	0.048
trios_kills	0.046
solo_top1	0.046
lifetime_kills	0.045
solo_kills	0.045
lifetime_wins	0.045
lifetime_winratio	0.044
duos_winratio	0.044
trios_top3	0.044
trios_top6	0.039
solo_scorepermin	0.039
lifetime_score	0.039
duos_top1	0.038
duos_kills	0.038
trios_score	0.038
lifetime_top3/5/10	0.037
solo_score	0.036
trios_minutesplayed	0.036
lifetime_minutesplayed	0.035
lifetime_kpm	0.034
solo_top10	0.034
lifetime_scorepermin	0.033
lifetime_top6/12/25	0.032

# Box Plots for Match Statistics

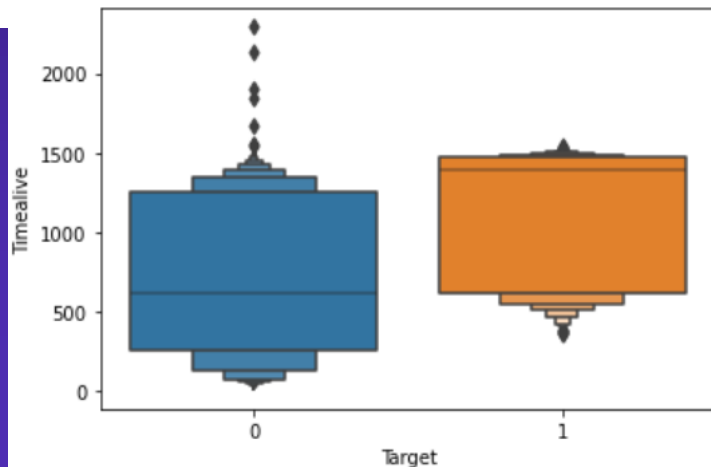
Distribution of eliminations



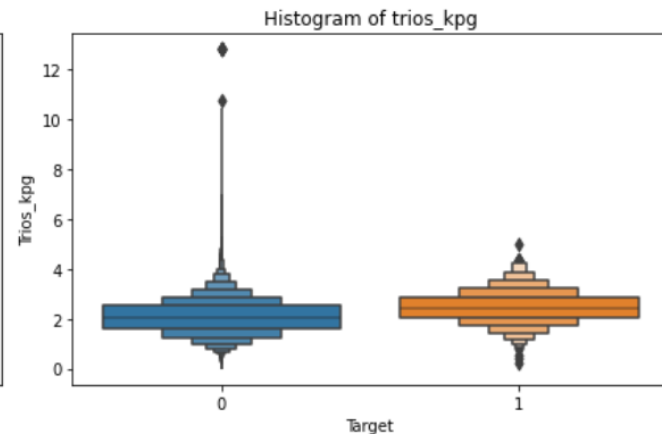
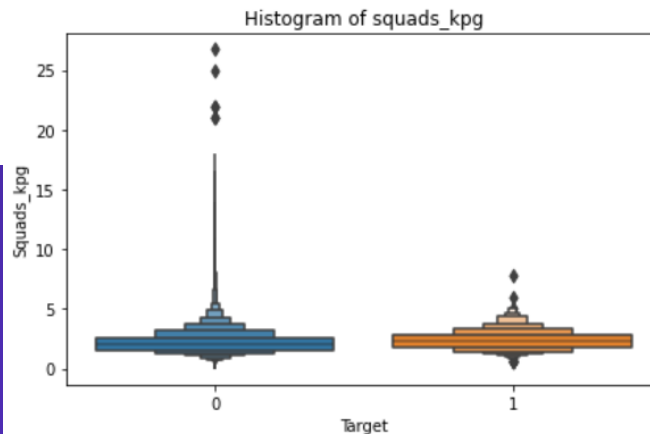
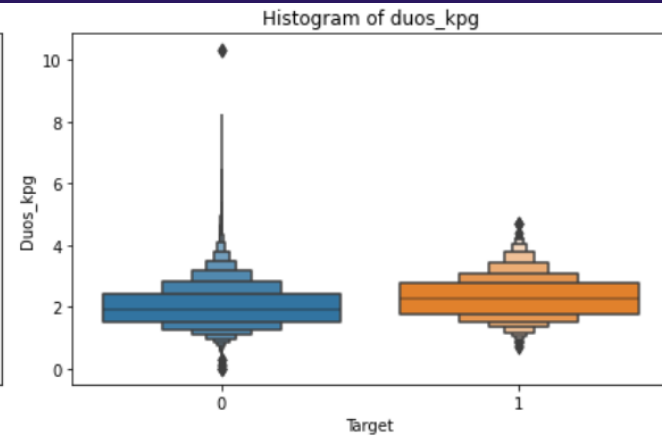
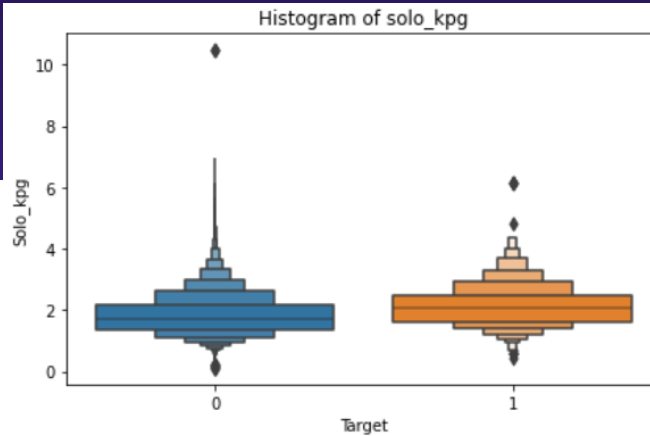
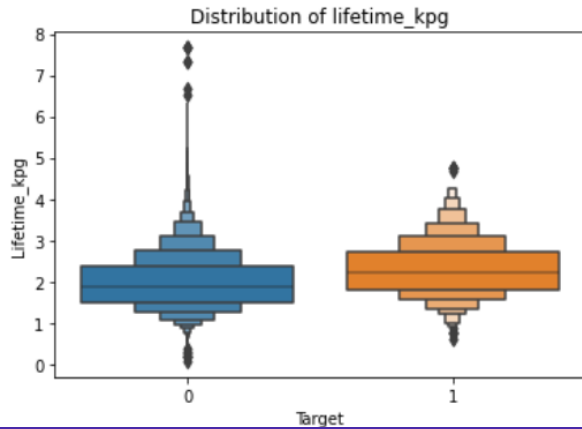
Distribution of points



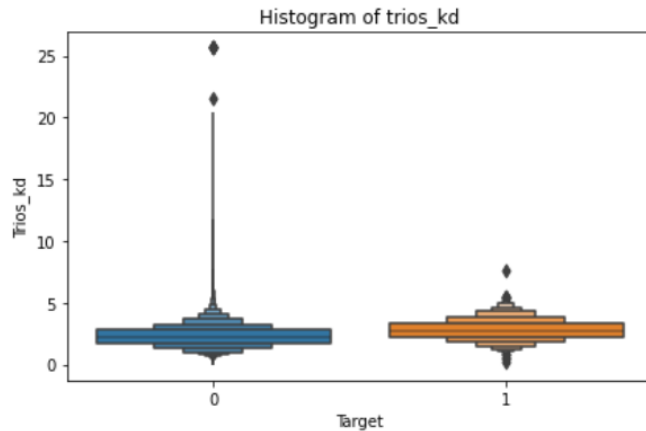
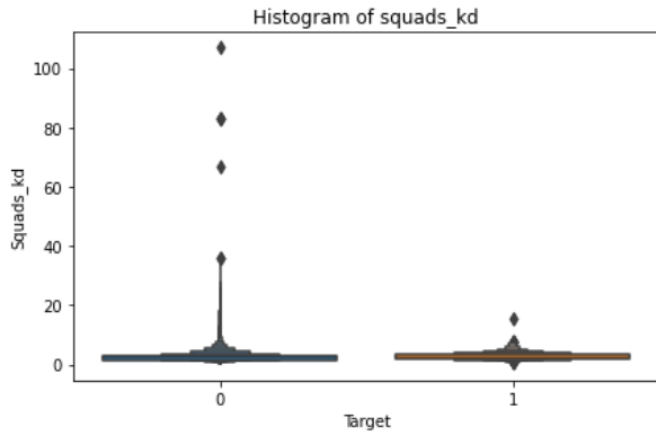
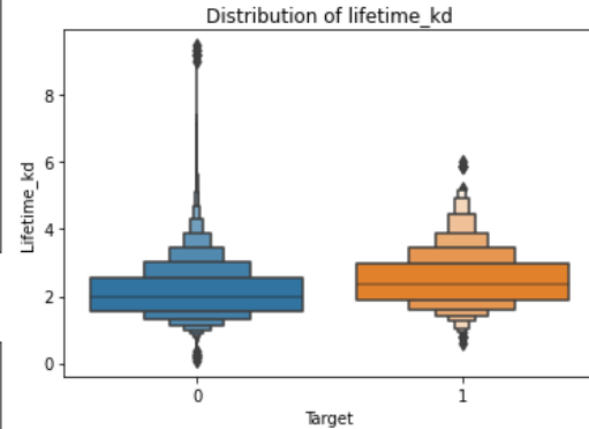
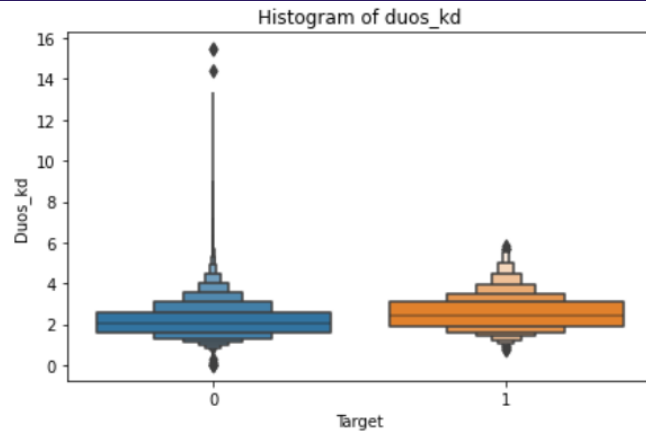
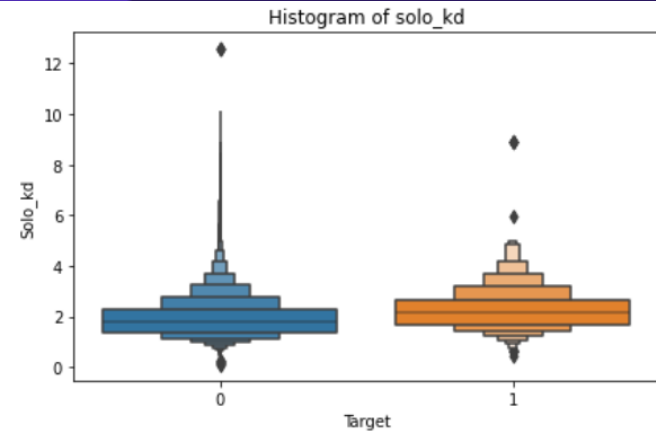
Distribution of timealive



# Box Plots for Player Statistics



# Box Plots for Player Statistics

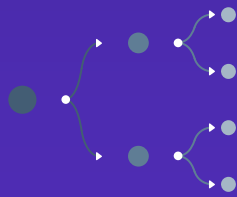




x x x x x  
x x x x x  
x x x x x  
x x x x x

04

Models



# Model Framework



## Basic Classifiers

- Logistics Regression
- K Neighbors Classifier



## Boost Classifier

- Light GBM Classifier



## Tree-based Classifier

- Random Forest Classifier



## Neural Network

- Keras Classifier

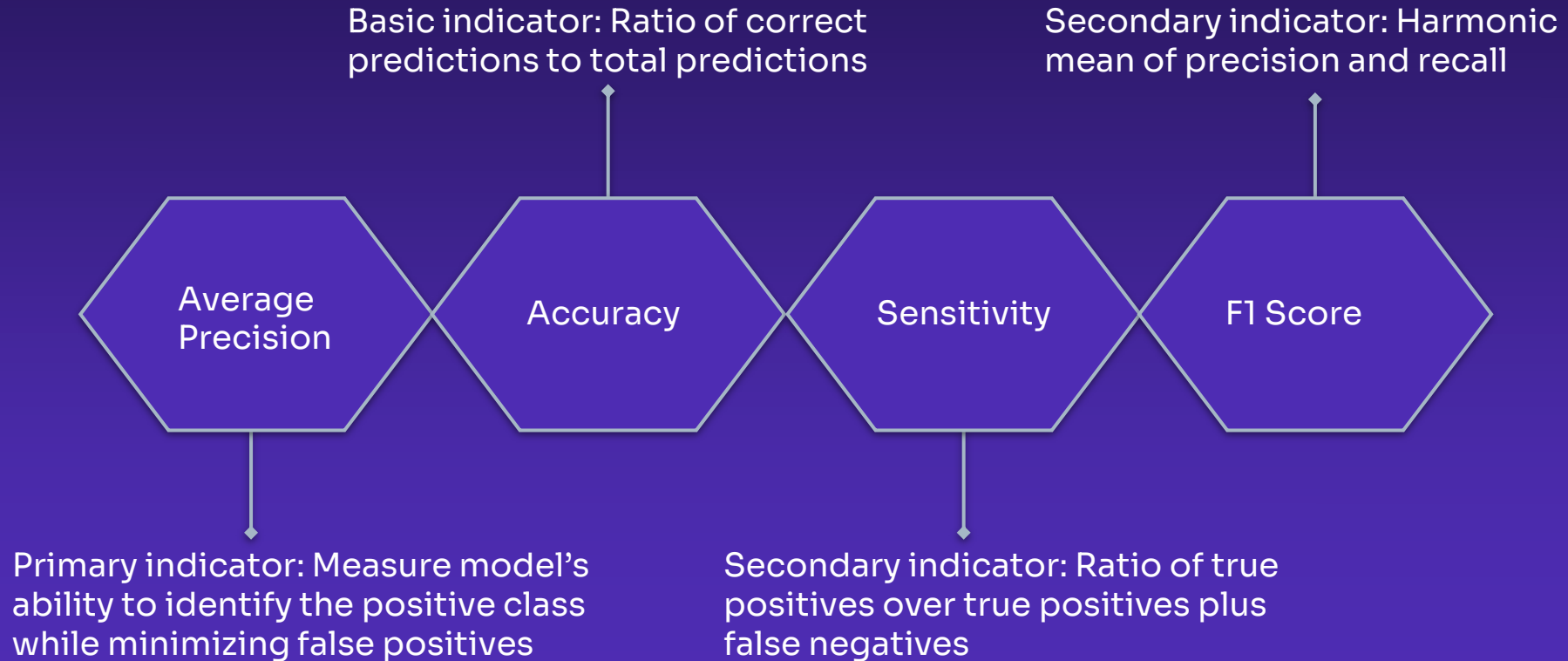
# Modelling

- Various parameters are used to tune the above models through the use of Grid Search CV.
- The results are as follows:

model	GridSearch_score	train_score	test_score	accuracy	recall	precision	roc_auc_score	f1score	average_precision
Dummy	0.0144	0.0144	0.0144	0.9856	0.0000	0.0000	0.5000	0.0000	0.0144
LogReg	0.0327	0.0347	0.0246	0.6611	0.5839	0.0247	0.6230	0.0473	0.0204
KNN	0.1520	0.2769	0.1559	0.6201	0.6510	0.0244	0.6353	0.0471	0.0209
RandomForest	0.1945	0.3222	0.1861	0.8310	0.4899	0.0419	0.6629	0.0771	0.0279
Light GBM	0.2177	0.4424	0.2230	0.9275	0.4497	0.0913	0.6921	0.1518	0.0490
Neural Network	0.1543	0.2896	0.2005	0.7551	0.5906	0.0344	0.6741	0.0650	0.0262

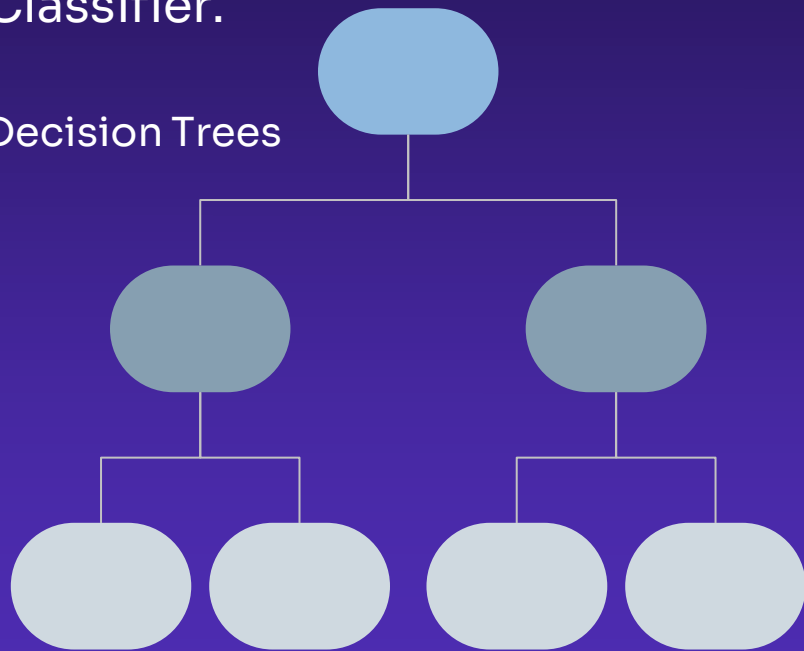


# Scoring Metrics



# Final Model

- The best model is the Light GBM Classifier.
- The parameters are as follows:
  - Boosting Type: Gradient Boosting Decision Trees
  - Scale Pos Weight: 99
  - Max Bin: 200
  - N Estimators: 200
  - Learning Rate: 0.01
  - Max Depth: 25
  - Num Leaves: 255
  - Min Child Samples: 200
  - Colsample by Tree: 0.9
  - Subsample: 0.9
  - Subsample Freq: 2



xx xx xx xx  
xx xx xx xx  
xx xx xx xx  
xx xx xx xx

# 05

## Evaluation



# Confusion Matrix

Predicted  
Non-Winners

Predicted  
Winners

Actual Non-  
Winners

Actual Winners

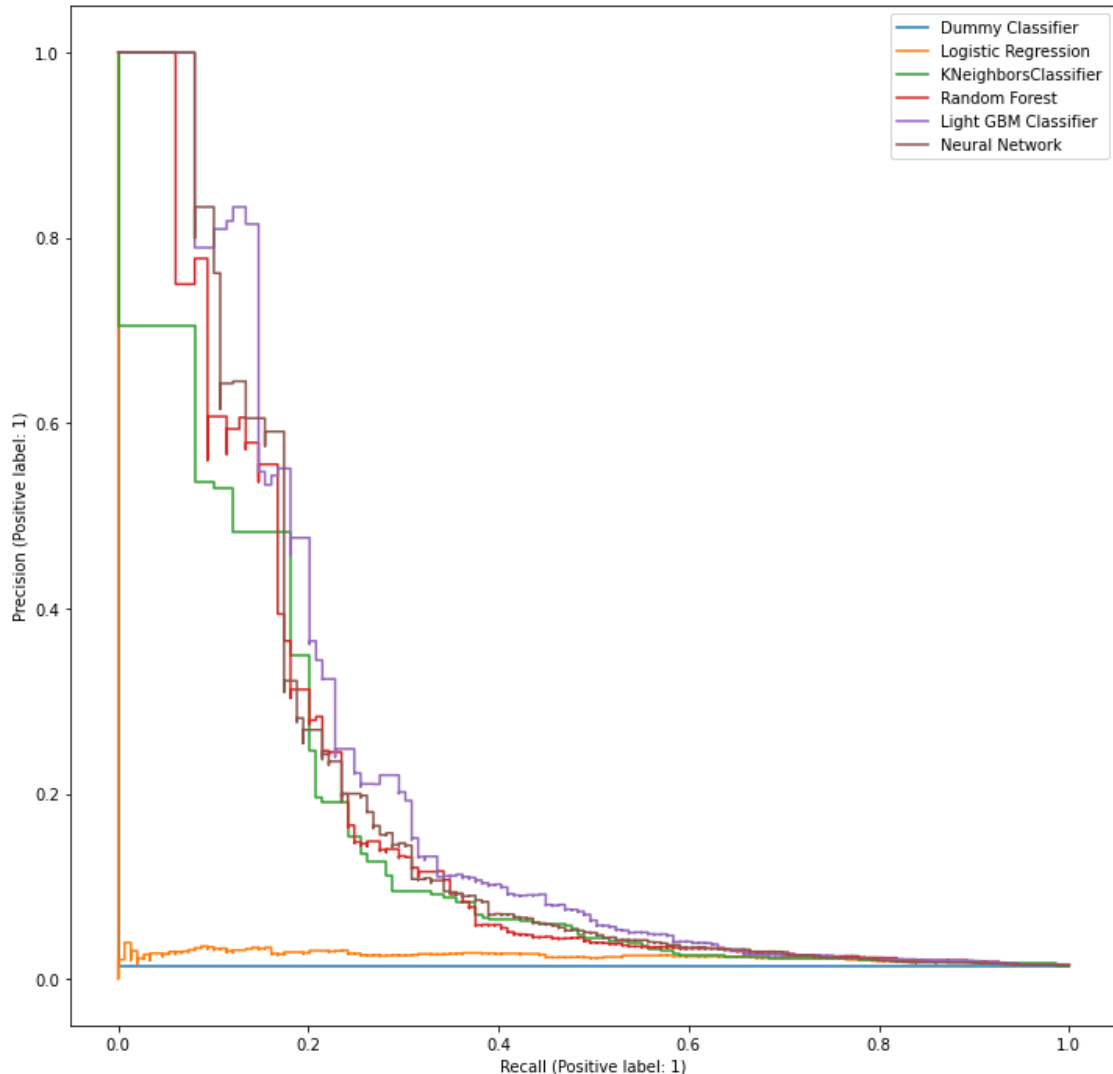
9,519	667
82	67

# Precision-Recall Curve

**Average  
Precision:**

0.0490,

Compared to  
Baseline of  
0.0144



# Other Scoring Metrics



**Accuracy:**

0.9275,

Compared to  
Baseline of  
0.9856



**Sensitivity:**

0.4497,

Compared to  
Baseline of 0



**F1 Score:**

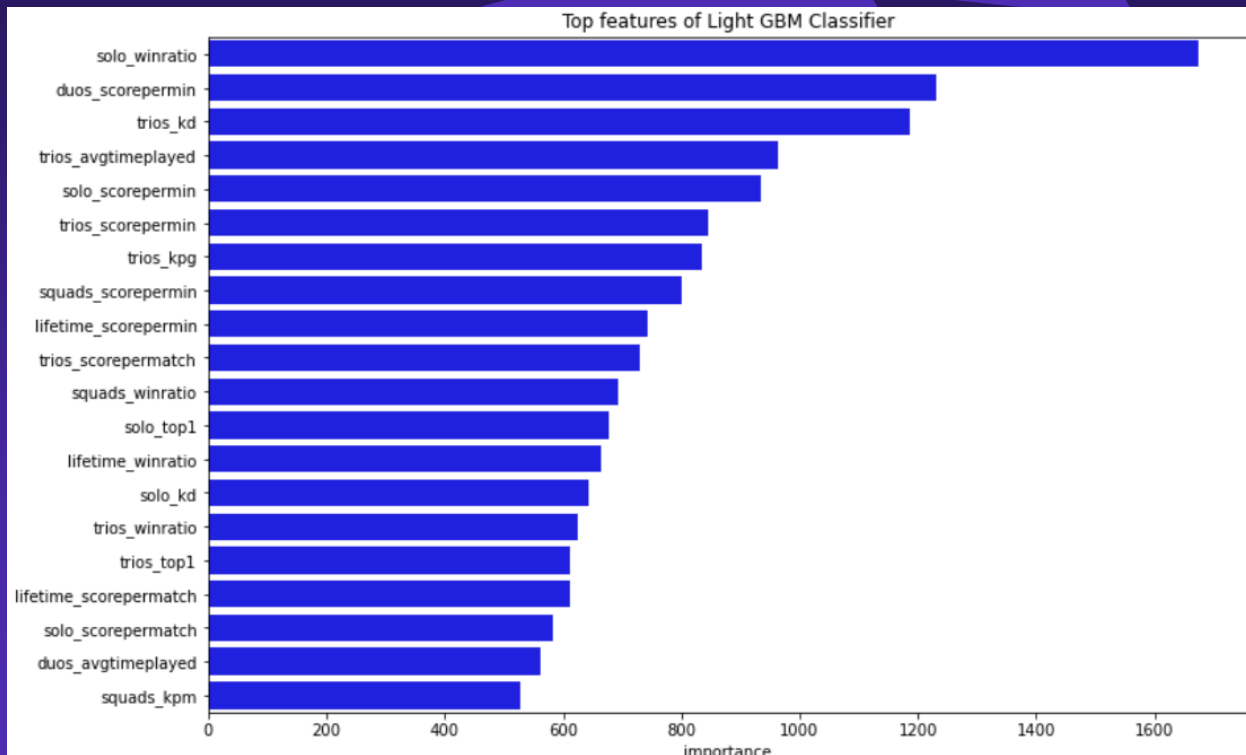
0.1518

Compared to  
Baseline of 0

# Recommendations

## Features to use to judge:

- Solo Win Ratio
- Score per min (for various match modes)
- Win Ratio (for various match modes)
- Ratio Statistics related to Trios match mode



Feature Importances of Final Production Model

# 06

## Conclusion





# Conclusion

## Improvements to current model:

- More match and player statistics data => Limited scale due to limitations of web scraping
- Other forms of player statistics not directly related to game-based => Not readily available and may be irrelevant

## Using match statistics to model:

- Model using match statistics like `eliminations` and `time alive` => Instead of predicting the winner, identify factors to increase chances of winning.

## Future Applications to other games

# Thanks!

Do you have any questions?

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution

