

# League of Legends



**Integrantes:**  
• Santiago Basañes  
• Santiago Cifuentes  
• Carolina Vlatko



**Laboratorio de Datos. 1er cuatrimestre 2021**  
**Cátedra Prof. Enzo Tagliazucchi**

# Introducción



# ¿Qué es el



**League of Legends es uno de los juegos de estrategia más jugados del mundo, con más de 100 millones de usuarios activos.**

Es aparte uno de los e-sport más vistos en Twitch, y la final de 2020 fue el evento e-sport más visto del año.



# ¿Qué es el



# ?



League of Legends es uno de los juegos de estrategia más jugados del mundo, con más de 100 millones de usuarios activos.

**Es aparte uno de los e-sport más vistos en Twitch, y la final de 2020 fue el evento e-sport más visto del año.**



# ¿Cómo se juega?

Dos equipos de 5 jugadores se enfrentan en un mapa casi simétrico con el objetivo de destruir el “NEXO” enemigo.



Se elige un personaje entre más de 140 campeones.

# ¿Cómo se juega?

Dos equipos de 5 jugadores se enfrentan en un mapa casi simétrico con el objetivo de destruir el “**NEXO**” enemigo.



Se elige un personaje entre más de 140 campeones.

**La partida se gana tras destruir una serie de objetivos**



# Hipótesis y preguntas



Los jugadores tiene ciertas **creencias** sobre qué campeones, qué objetivos y cuáles decisiones son las correctas, y estas varían de acuerdo a su nivel de habilidad.

 Posted by u/StuntinOutFront   Gunblade Gang 12 days ago 

**867 Is Akali broken in pro play??**

 After watching pro games in the last few weeks, from all the major regions, I have seen Akali pop off too many games already. Although her early game is weaker with the mini rework, her mid-late game is strong. Is she overbuffed now? I anticipate some damage reduction in her E over the next few patches.

 Posted by u/This\_Hornet\_3367 15 days ago   

**1.4k Ezreal is currently completely broken**

 Ezreal has almost 52% WR in plat+ with almost 40% PR. He is considered to be balanced on around 48%, as he is hard, skillshot-heavy champion with high pick ratio. Recent buffs of Sunderer and the champion himself made Ezreal completely unbalanced. I have no idea how Kai'Sa and Varus were hit by nerfs, while Ezreal dodged them, making his winrate skyrocket to even higher amounts.

<https://u.gg/lol/champions/ezreal/build>



Los jugadores tiene ciertas **creencias** sobre qué campeones, qué objetivos y cuáles decisiones son las correctas, y estas varían de acuerdo a su nivel de habilidad.

 Posted by u/StuntinOutFront   Gunblade Gang 12 days ago 

### 867 Is Akali broken in pro play??



After watching pro games in the last few weeks, from all the major regions, I have seen Akali pop off too many games already. Although her early game is weaker with the mini rework, her mid-late game is strong. Is she overbuffed now? I anticipate some damage reduction in her E over the next few patches.

 Posted by u/This\_Hornet\_3367 15 days ago   

### 1.4k Ezreal is currently completely broken



Ezreal has almost 52% WR in plat+ with almost 40% PR. He is considered to be balanced on around 48%, as he is hard, skillshot-heavy champion with high pick ratio. Recent buffs of Sunderer and the champion himself made Ezreal completely unbalanced. I have no idea how Kai'Sa and Varus were hit by nerfs, while Ezreal dodged them, making his winrate skyrocket to even higher amounts.

<https://u.gg/lol/champions/ezreal/build>

**Desarrollamos modelos para responder a preguntas específicas sobre el balance del juego y los factores que determinan la victoria**



# Hipótesis

- **Basta conocer los campeones de una partida para saber qué equipo ganará la partida.**
- **Basta conocer los campeones y *compañeros (jugadores)* para saber qué equipo ganará la partida:** ¿Es el *dodgeo* es una decisión justificada?
- **Los primeros objetivos neutrales definen la partida.**
- **A mayor liga, más precisos son los modelos:** ¿El *meta* tiene menos impacto en las ligas menores?



# Armando dataset: API RIOT.



ApiRiot: <https://developer.riotgames.com/>

**Pro:** Nos permite armar nuestro dataset.

**Contra:** Tiene **límites restrictivos** de consultas por segundo y minuto.



ApiRiot: <https://developer.riotgames.com/>

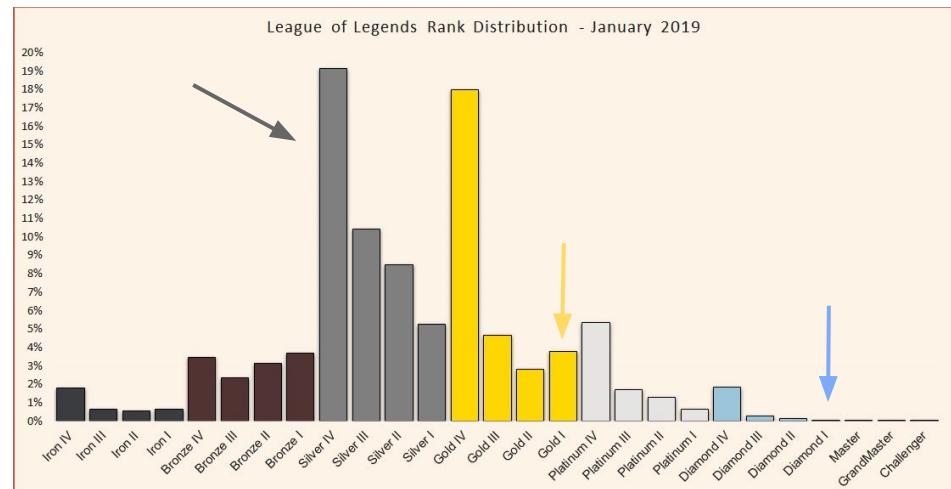


**Pro:** Nos permite armar nuestro dataset.

**Contra:** Tiene **límites restrictivos** de consultas por segundo y minuto.

Bajamos los **resúmenes** de ~15 mil **partidas** de tres divisiones:

- Silver IV
- Gold I
- Diamond I



# Modelo I: Relevancia de los campeones



# El Balance

Riot se esfuerza por mantener el juego “**balanceado**”.



**11.8**  
NOTAS DE LA VERSIÓN

▼ NERFS

- Gnar
- Annie
- Orianna
- Thresh
- Yorick

▲ MEJORAS

- Aphelios
- Cassiopeia
- LeBlanc
- Lee Sin
- Rammus
- Vladimir
- Zac

Ningún campeón debe ser mejor que el resto.  
De esta forma todo jugador puede usar el campeón que quiere sin ver su performance gravemente afectada.

# El Balance

Riot se esfuerza por mantener el juego “**balanceado**”.



**L 11.8 NOTAS DE LA VERSIÓN**

▼ NERFS

- Gnar
- Annie
- Orianna
- Thresh
- Yorick

▲ MEJORAS

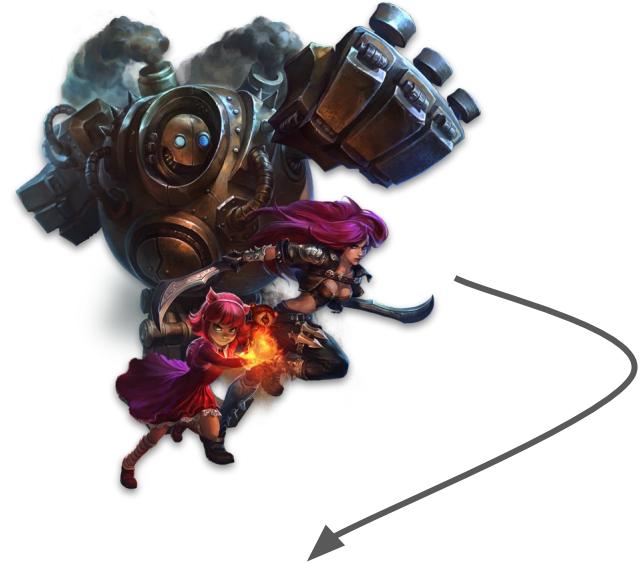
- Aphelios
- Cassiopeia
- LeBlanc
- Lee Sin
- Rammus
- Vladimir
- Zac

Ningún campeón debe ser mejor que el resto. De esta forma todo jugador puede usar el campeón que quiere sin ver su performance gravemente afectada.

**No se debería poder deducir el resultado de la partida a partir de los campeones**

# Representando un equipo

Podemos usar un vector de 155 posiciones, donde cada posición se corresponde con un personaje



Annie	Blitz	Kata	Teemo
1	1	1	0

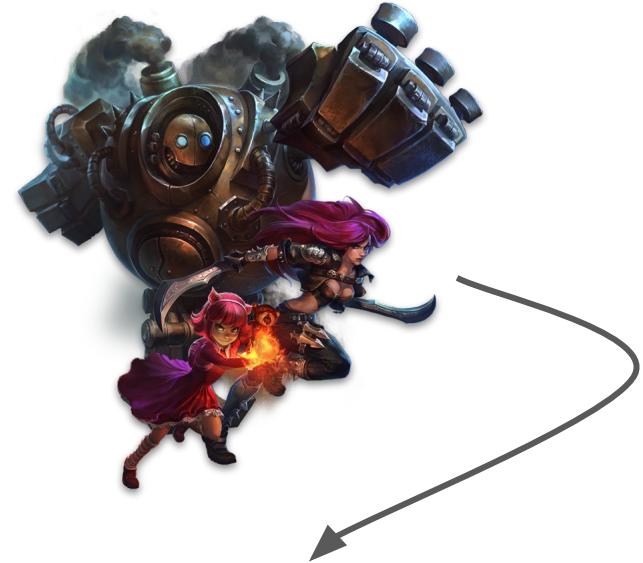
# Representando un equipo

Podemos usar un vector de 155 posiciones, donde cada posición se corresponde con un personaje

**Pros:** Cada campeón está representado en su totalidad.



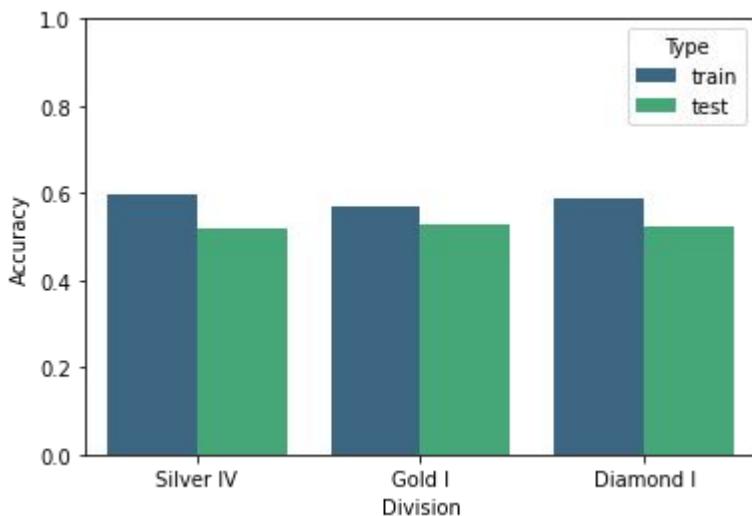
**Contra:** Son 310 features, y estos no representan ninguna interacción entre personajes aliados o enemigos.



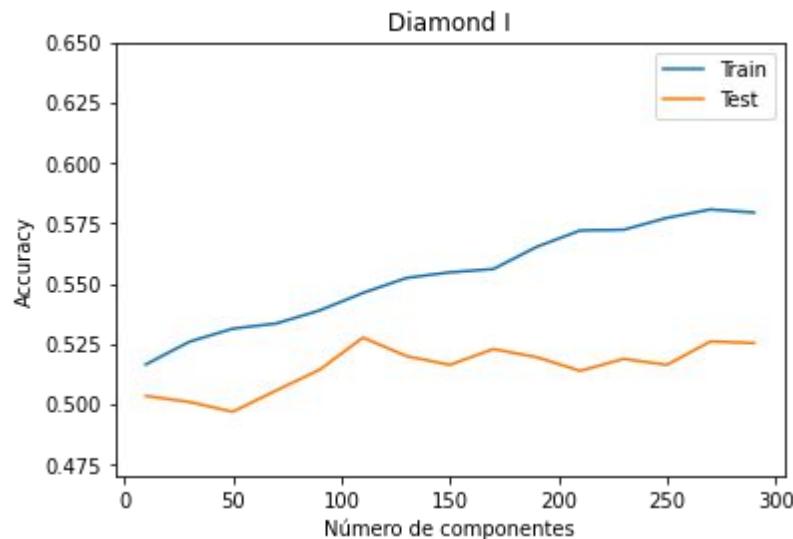
Annie	Blitz	Kata	Teemo
1	1	1	0

# Resultados

## Regresión logística sin regularizar

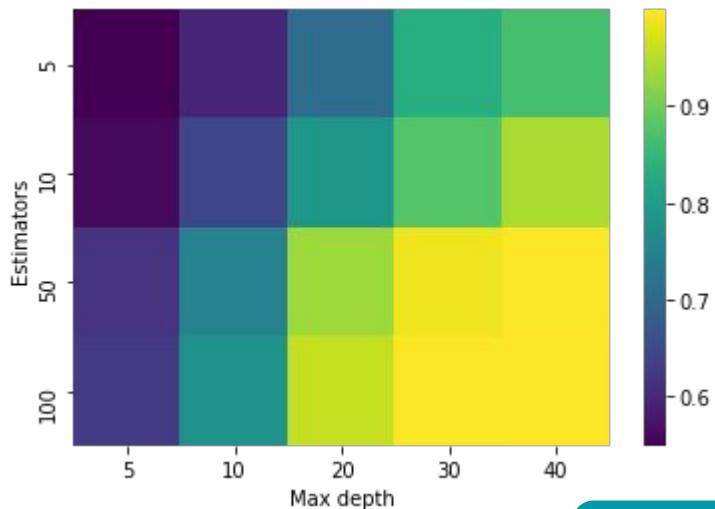


## Regresión logística con componentes de PCA

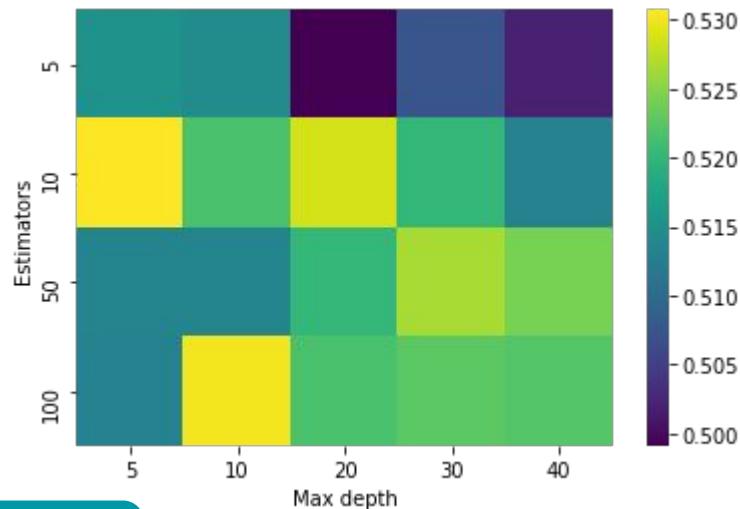


# Resultados

Accuracy en train para distintos modelos de RandomForest



Accuracy en test para distintos modelos de RandomForest



Por lo menos fittea...

# Resultados

- No parece que haya nada bueno que generalizar a partir de estas representaciones



# Resultados

- No parece que haya nada bueno que generalizar a partir de estas representaciones
- Puede deberse a que no hay relaciones marcadas entre los campeones elegidos y la victoria



# Resultados

- No parece que haya nada bueno que generalizar a partir de estas representaciones
- Puede deberse a que no hay relaciones marcadas entre los campeones elegidos y la victoria
- O puede que con una representación mejor se obtengan resultados positivos, pero...

## Es difícil “clusterizar” campeones

Las partidas ranked siguen templates básicos, y es muy fácil perder rápido las diferencias entre partidas

## Es difícil combinar atributos

Solo contando interacciones a pares ya hay más de  $10^5$  features. Analizar con PCA tampoco ayudó...



# Modelo II

## Repensando el dataset



**¿Cómo podemos generar un dataset que contenga información que caracterice a los campeones sin ir poniendo todo los campeones posibles?**





CHAMPION  
**MATCH UP**



Twisted Fate  
50.72%

VS

Fizz  
49.28%

Enfrentamiento

**26.82%**

Probabilidad de Lane Kill

Enfrentamiento

**73.18%**

**2.07 : 1**

KDA

**2.84 : 1**

**43.9%**

Contribución a las Kills

**47.3%**

**13,601**

Daño infligido a campeones

**17,100**

**17'58"**

Primera torreta destruida

**17'52"**

**50.72%**

Tasa de victoria

**49.28%**

**49.69%**

Tasa de victoria según la posición

**48.46%**

**6.82%**

Tasa de selección según la posición

**2.82%**

**1.05%**

Tasa/Ban

**4.53%**

Oro por minuto

CS por minuto

Experiencia por minuto

450

400

350

300

0-10

10-20

20-30

30+

# Matriz de winrate

ID Campeón rival

cid	matchs	121	122	126	127	131	133
121	0.07	0.00	0.41	0.50	0.44	0.43	0.18
122	0.08	0.59	0.00	0.31	0.71	0.45	0.70
126	0.02	0.50	0.69	0.00	0.75	0.47	0.40
127	0.01	0.56	0.29	0.25	0.00	0.12	0.00
131	0.12	0.57	0.55	0.53	0.88	0.00	0.31
133	0.02	0.82	0.30	0.60	1.00	0.69	0.00

Diagrama de la Matriz de winrate:

- ID Campeón:** Indica la fila de la matriz.
- ID Campeón rival:** Indica la columna de la matriz.
- % de victorias:** Indica el porcentaje de victorias para cada par de campeones.
- %Partidas en el dataset:** Indica el porcentaje de partidas en el dataset para cada par de campeones.

Notas:

- Los valores en la diagonal principal (121, 122, 126, 131, 133) representan el % de victorias del campeón contra sí mismo.
- Los valores en la fila 127 (0.01, 0.12, 0.00, 0.00) representan el % de victorias del campeón 127 contra los demás campeones.
- Los valores en la fila 131 (0.12, 0.57, 0.00, 0.31) representan el % de victorias del campeón 131 contra los demás campeones.

ID Campeón

%Partidas  
en el dataset

# Matriz de winrate



## Rivales-línea

cid	matchs	121	122	126	127	131	133
121	0.07	0.00	0.41	0.50	0.44	0.43	0.18
122	0.08	0.59	0.00	0.31	0.71	0.45	0.70
126	0.02	0.50	0.69	0.00	0.75	0.47	0.40
127	0.01	0.56	0.29	0.25	0.00	0.12	0.00
131	0.12	0.57	0.55	0.53	0.88	0.00	0.31
133	0.02	0.82	0.30	0.60	1.00	0.69	0.00

## Rivales-general

cid	matchs	121	122	126	127	131	133
121	0.07	0.00	0.41	0.50	0.44	0.43	0.18
122	0.08	0.59	0.00	0.31	0.71	0.45	0.70
126	0.02	0.50	0.69	0.00	0.75	0.47	0.40
127	0.01	0.56	0.29	0.25	0.00	0.12	0.00
131	0.12	0.57	0.55	0.53	0.88	0.00	0.31
133	0.02	0.82	0.30	0.60	1.00	0.69	0.00

## Compañeros

cid	matchs	121	122	126	127	131	133
121	0.07	0.00	0.41	0.54	0.14	0.59	0.50
122	0.08	0.41	0.00	0.60	0.38	0.47	1.00
126	0.02	0.54	0.60	0.00	0.50	0.50	0.50
127	0.01	0.14	0.38	0.50	0.00	0.67	1.00
131	0.12	0.59	0.47	0.50	0.67	0.00	0.58
133	0.02	0.50	1.00	0.50	1.00	0.58	0.00

# Dataset

top1_winrate_lane	top1_winrate_partner_{i} (x4)	top1_winrate_rival_{i} (x4)	top1_main_role	top1_matchs
0.3	0.4	0.1	0	0.04

top2_winrate_lane	top2_winrate_partner_{i} (x4)	top2_winrate_rival_{i} (x4)	top2_main_role	top2_matchs
0.7	0.2	0.33	1	0.08

# Dataset

top1_winrate_lane	top1_winrate_partner_{i} (x4)	top1_winrate_rival_{i} (x4)	top1_main_role	top1_matchs
0.3	0.4	0.1	0	0.04

top2_winrate_lane	top2_winrate_partner_{i} (x4)	top2_winrate_rival_{i} (x4)	top2_main_role	top2_matchs
0.7	0.2	0.33	1	0.08

**+Jungla, +Mid, +Support, +Bot**

**Pros:** Se representa cada campeón por su winrate en relación con los otros campeones.



**Pros:** El impacto de cada uno de las líneas/roles queda representado dentro del data set.



**Pros:** Son la mitad de features ( $10\text{champs} \times (\text{wr\_lane} + \text{wr\_partner} \times 4 + \text{wr\_rivals} \times 4 + \text{wr\_role} + \text{wr\_size}) + \text{team\_read}$ ) = 111 features

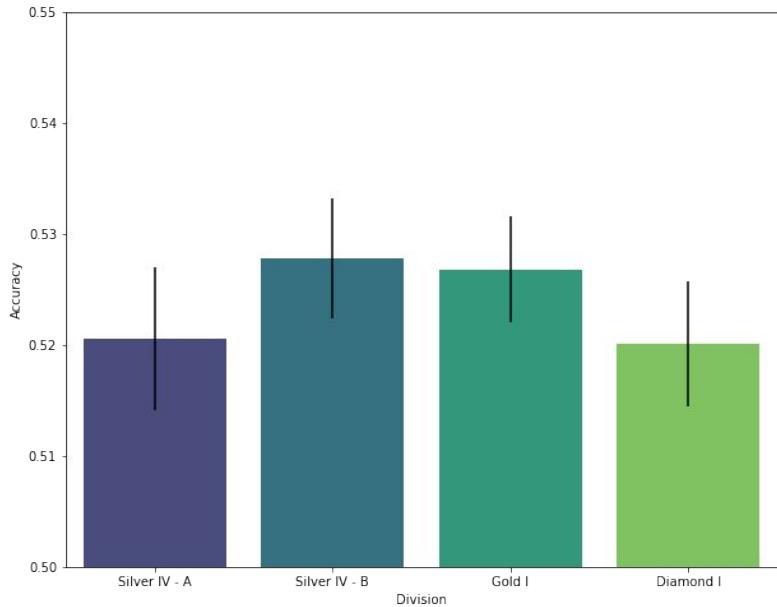
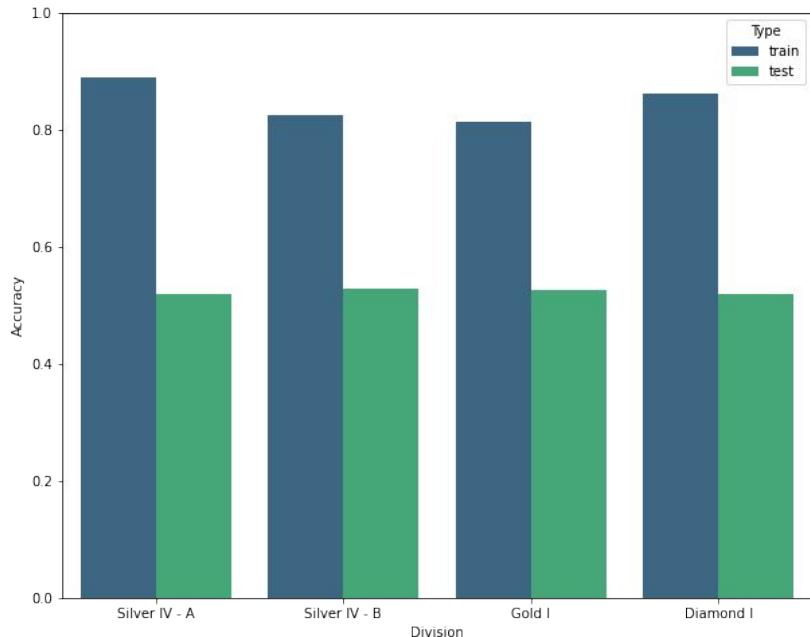


**Contra:** El tiempo de creación del nuevo dataset es de +30' (depende de la muestra de partidas que se tenga).



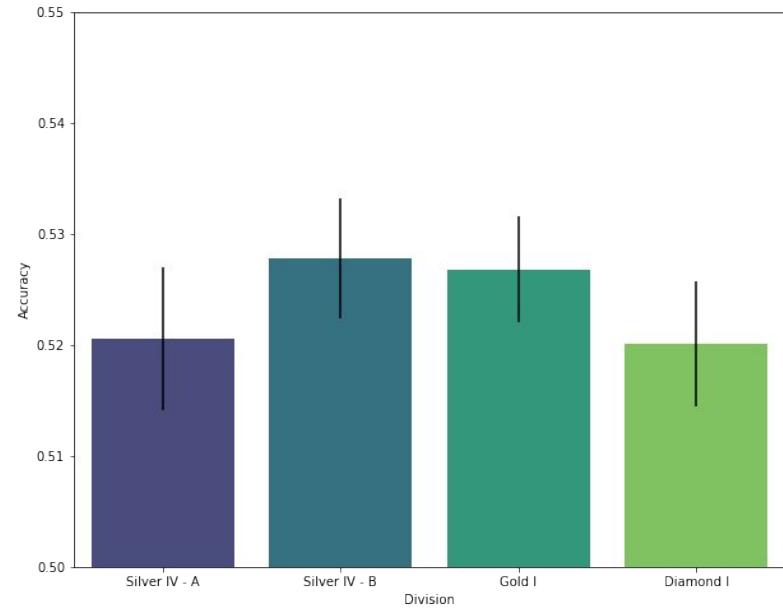
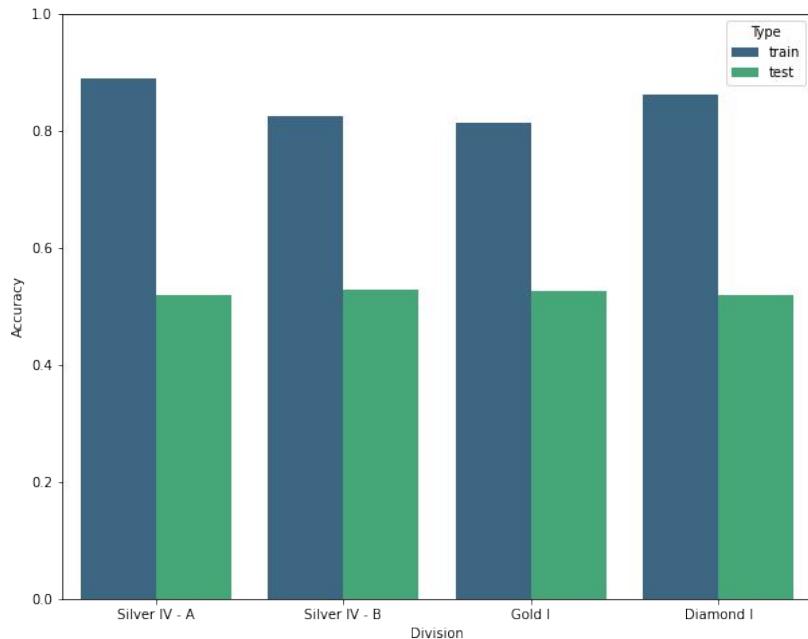
# Resultados

## Regresión logística



# Resultados

## Regresión logística



**Se probaron varias estrategias (valores medios, interacciones entre algunos features, reemplazar el *winrate* por su distancia al valor medio, etc)**

# Resultados

El modelo sigue sin ser un buen predictor pero...



# Resultados

El modelo sigue sin ser un buen predictor pero...

## **Se mejora el accuracy del set de entrenamiento**

Muestra que tal vez este sea el camino indicado.

## **Aún hay métricas por agregar**

KDA medio, Daño partida, Dificultad de uso del campeón, etc

## **Una mejor construcción del dataset**

Armar un dataset que tenga un mínimo de partidas por campeón puede ser de utilidad.

## **Otros modelos**

Hasta ahora solo usamos regresiones lineales y árboles. Hay más para probar.

## **Otros resultados**

Otros trabajos en internet con data SOLO de los campeones muestra valores similares a los nuestros...

# Resultados

El modelo sigue sin ser un buen predictor pero...

## Se mejora el accuracy del set de entrenamiento

Muestra que tal vez este sea el camino indicado.

## Aún hay métricas por agregar

KDA medio, Daño partida, Dificultad de uso del campeón, etc

## Una mejor construcción del dataset

Armar un dataset que tenga un mínimo de partidas por campeón puede ser de utilidad.

## Otros modelos

Hasta ahora solo usamos regresiones lineales y árboles. Hay más para probar.

## Otros resultados

Otros trabajos en internet con data SOLO de los campeones muestra valores similares a los nuestros...

**Los campeones no marcan una diferencia para los jugadores casuales.**

# Resultados

El modelo sigue sin ser un buen predictor pero...

## **Se mejora el accuracy del set de entrenamiento**

Muestra que tal vez este sea el camino indicado.

## **Aún hay métricas por agregar**

KDA medio, Daño partida, Dificultad de uso del campeón, etc

## **Una mejor construcción del dataset**

Armar un dataset que tenga un mínimo de partidas por campeón puede ser de utilidad.

## **Otros modelos**

Hasta ahora solo usamos regresiones lineales y árboles. Hay más para probar.

## **Otros resultados**

Otros trabajos en internet con data SOLO de los campeones muestra valores similares a los nuestros...

**No se puede deducir el resultado de la partida a partir de los campeones**

# Modelo III: Objetivos del juego



# Objetivos del juego

Por cada uno tenemos: [First, kills]

Equipo

Campeones  
Torres  
Inhibidores

Neutrales

Heraldo  
Dragón  
Barón

*Con un modelo de regresión logística queremos predecir si un equipo gana según un subconjunto de objetivos cumplidos*

# Objetivos del juego

Por cada uno tenemos: [First, kills]

Equipo

Campeones  
Torres  
Inhibidores

Neutrales

Heraldo  
Dragón  
Barón

Juego temprano

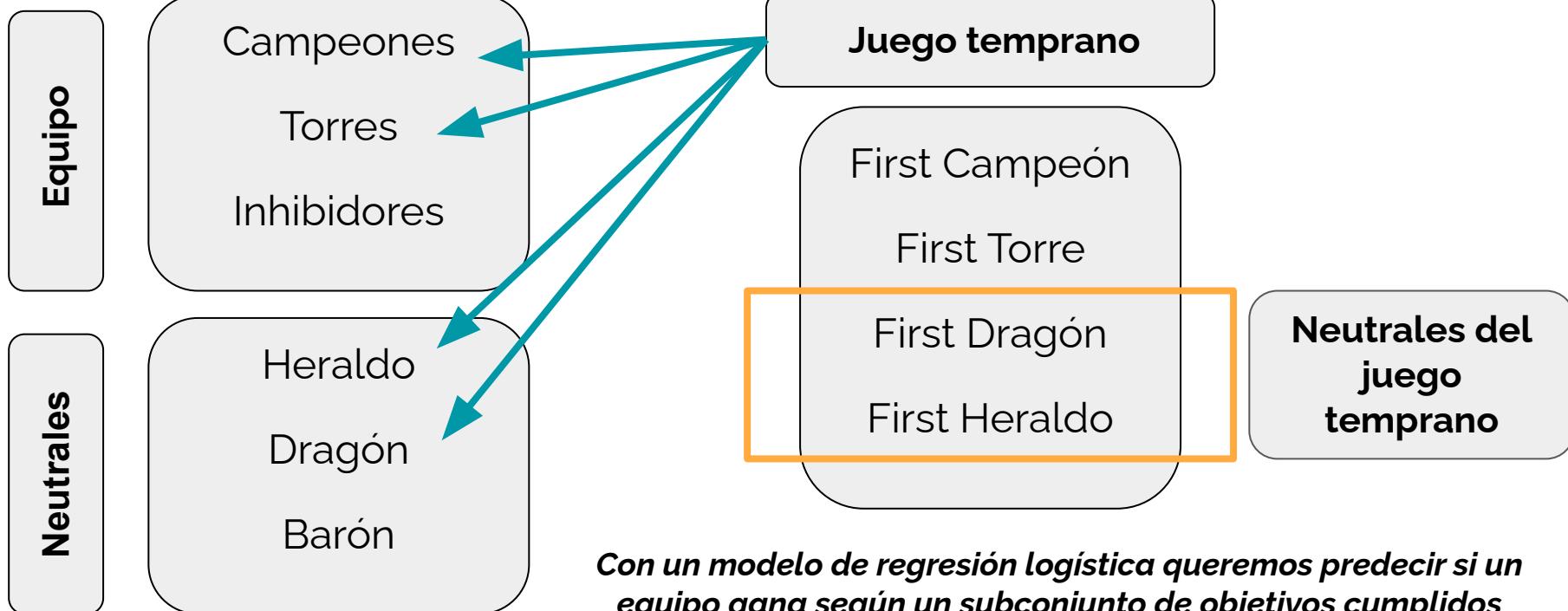
primeros ~10 min

First Campeón  
First Torre  
First Dragón  
First Heraldo

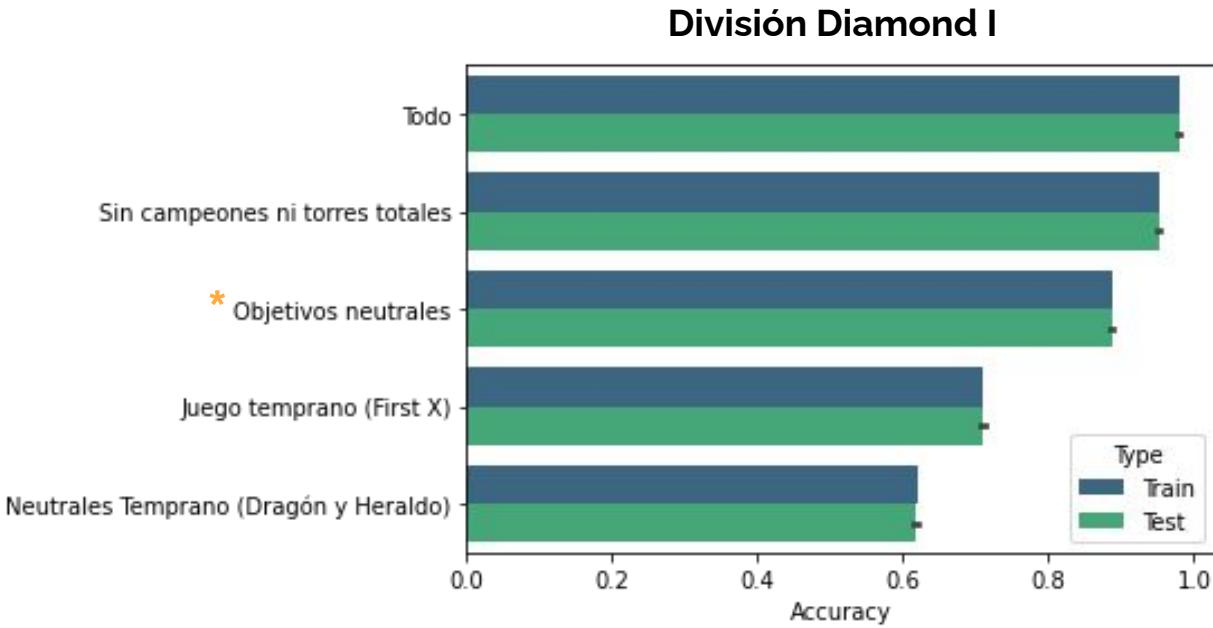
*Con un modelo de regresión logística queremos predecir si un equipo gana según un subconjunto de objetivos cumplidos*

# Objetivos del juego

Por cada uno tenemos: [First, kills]



# Regresión logística sin regularización



Acierto de 90% con sólo los **objetivos neutrales**.

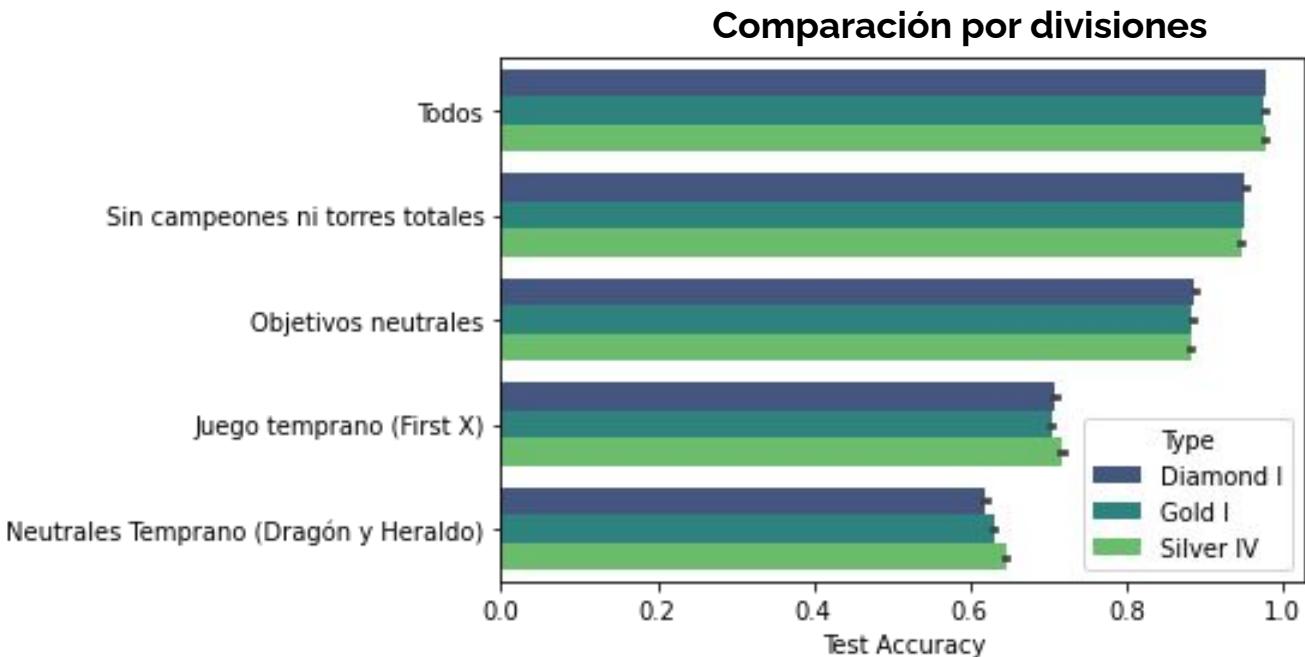
Solo con los primeros ~10 min. de partida (Juego temprano) el acierto es del **70%**.

Solo con el First Dragón y First Heraldo el modelo es mejor que tirar una moneda (61%).



\* barón >> dragón y Heraldo

# ¿Diferencia entre divisiones?



No vimos una diferencia significativa entre las divisiones, al contrario de lo que planteamos como **hipótesis**.

Sí vimos que en la "peor" división los **objetivos tempranos** tienen una correlación un **3% consistentemente** mayor con la victoria que la "mejor" división.

# Conclusiones





# Conclusiones

- El resultado de una partida no parece depender solo de los campeones. Se podría decir que el juego se decide una vez en el mapa, y no en champ select.
- Sin embargo los eventos al comienzo de la partida (~10 minutos) sí permiten tener una moderada seguridad sobre qué equipo va a ganar.
- También vimos que en todos los casos los objetivos neutrales tienen un gran impacto en el desenlace de la partida.
- No hubo diferencias considerables entre las distintas ligas analizadas.



# iGracias!

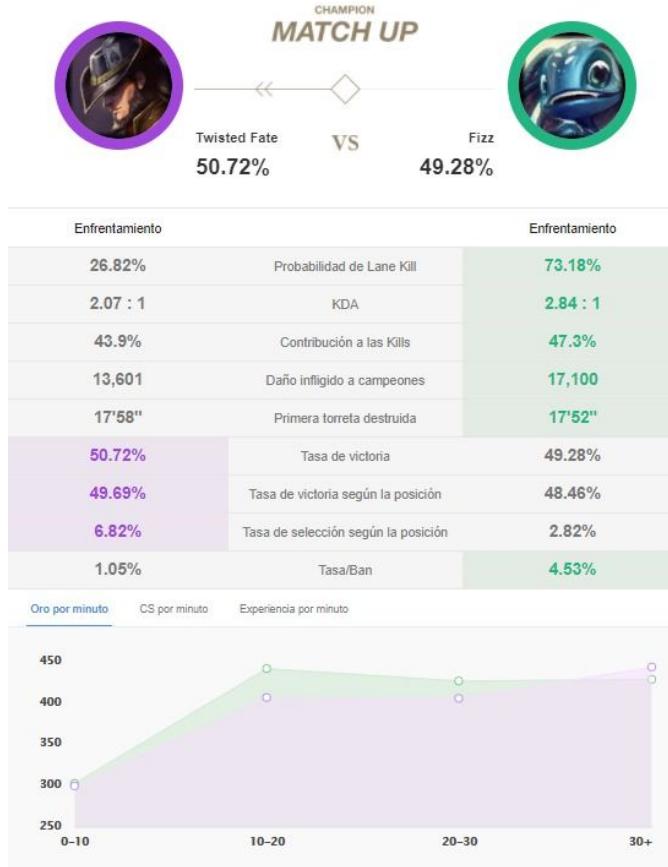




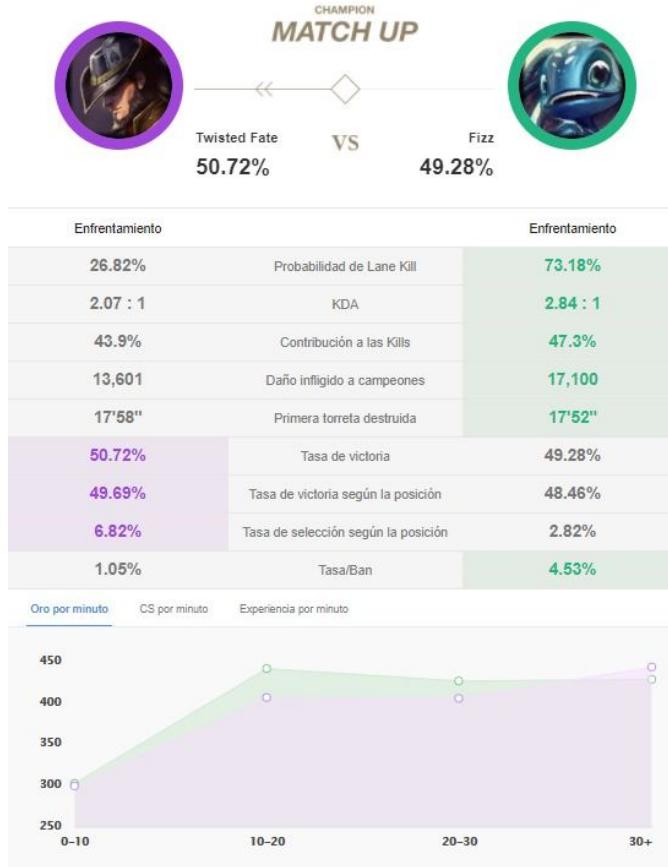
# Preguntas



Al ser un juego basado fundamentalmente en **continuas decisiones** (elección del campeón, itemización, maestrías, objetivos) hay varios sitios que se dedican a recopilar **estadísticas del juego** para ayudar a los jugadores.



Al ser un juego basado fundamentalmente en **continuas decisiones** (elección del campeón, itemización, maestrías, objetivos) hay varios sitios que se dedican a recopilar **estadísticas del juego** para ayudar a los jugadores.



**¿De donde sacan la data?**

# Índice

- **Introducción:** ¿Qué es el League of legends?
- **Armando dataset:** ApiRiot.
- **Hipótesis y preguntas.**
- **Modelo I:** Mirando campeones.
- **Modelo II:** Mirando estadísticas de las partidas.
- **Modelo III:** Los objetivos y su relevancia.
- **Conclusiones.**
- **¿Esto como sigue?**



