

Partidas profesionales de League of Legends correspondientes a de WORLDS, LEC y LCS 2020.

Manuel Ruiz Botella

9 de noviembre de 2020

Contexto.

Los datos recolectados corresponden a los datos de mapas de partidas de League of Legends entre jugadores profesionales en torneos oficiales, concretamente a los torneos WORLDS, LEC y LCS de 2020. Se han tratado todos los mapas jugados en estos torneos a lo largo de todo el año 2020, incluyendo información tanto de las fases regulares de primera y verano como de los playoffs. Respecto al mundial se ha incluido tanto los mapas de la fase de play-in como de la fase general y playoff.

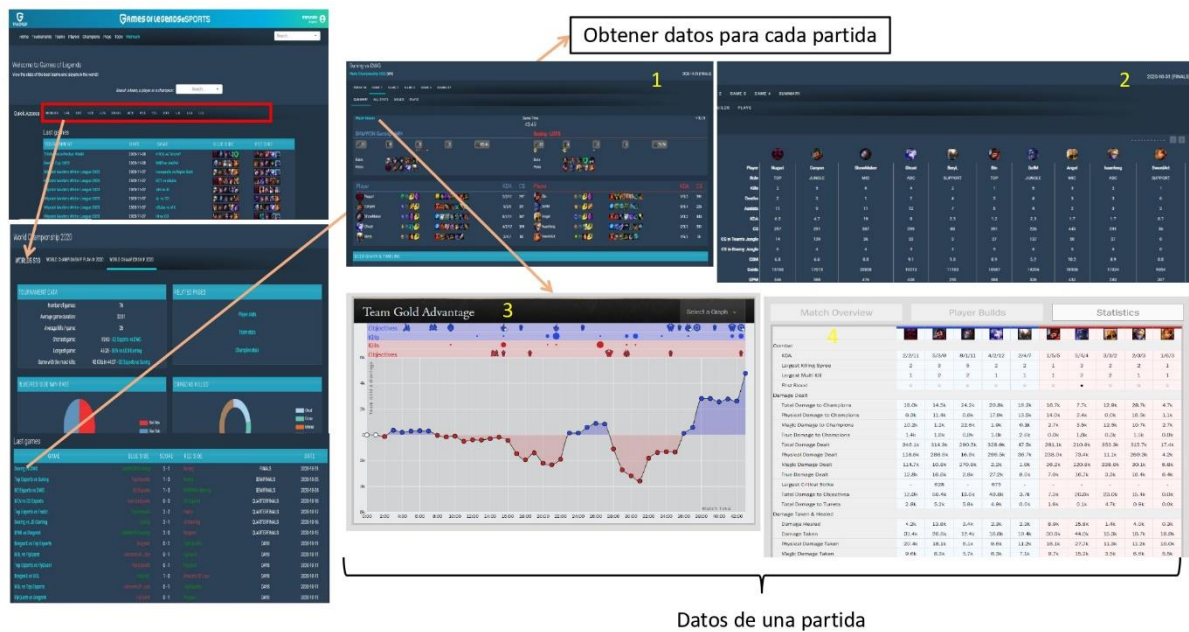
El sitio web principal utilizado gol.gg, es un sitio referente en base a estadísticas de partidas profesionales de League of Legends utilizado por analistas para estudiar las competiciones, por tanto, proporcionan una gran cantidad de información por cada mapa de una partida de League of Legends. Por otra parte, se accede también a leagueoflegends.com, que es la página web oficial del videojuego, y por tanto proporcionan la información de la partida porque obviamente son ellos quien posee la información.

Descripción.

Como ya se ha mencionado, el conjunto de datos se corresponde a **los datos que se generan en cada mapa de una partida de League of Legends jugada por jugadores profesionales, el dataset generado solo abarca las partidas de WORLDS, LEC y LCS de 2020**. Aunque el código está listo para acceder a cualquier otra liga importante como LCK o LPL.

En cada partida se recogen todo tipo de variables, desde el resultado del mapa, los equipos que lo juegan, los campeones utilizados, los jugadores de cada equipo, una gran cantidad de estadísticas del juego, etc. Algunas de las variables son básicas como el Oro final de un jugador, y otras derivadas como el oro por minuto. En total hay más de 500 variables, entre aquellas que son específicas de jugadores y aquellas que hacen referencia a un evento de la partida como la primera sangre o el número de dragones. Además, con las variables actuales es posible generar más variables al procesar los datos en función del objetivo.

Representación gráfica.



Contenido.

Los datos de cada mapa de una partida se recogen de 5 formas diferentes. Cada línea del dataset representa un mapa de una partida jugada entre dos equipos profesionales de League of Legends. Cada fila tiene más de 500 variables, pero muchas de ellas son la misma variable para cada jugador del mapa, por ejemplo, el número de asesinatos por jugador se divide en 10 variables en el dataset ya que tenemos 5 jugadores en 2 equipos. Las variables que tienen un valor para cada jugador en partida las explicaré solo una vez, no 10 veces.

Respecto al tiempo de los datos, el dataset corresponde solo los torneos de LEC, WORLDS y LCS de 2020, por lo que son datos de todo 2020, no entra en otros años.

El orden de cómo se obtienen los datos del dataset es:

1. Se recogen unos pocos datos de la página resumen de cada parte del torneo.
2. Para cada mapa de una partida entre dos equipos (una partida puede tener entre 1 y 5 mapas):
 - 2.1. Se recogen los datos accediendo al historial de partida de leagueoflegends.com.
 - 2.1.1. Se interactúa con unas gráficas de oro para obtener información de la diferencia de oro, el oro de cada equipo y el de oro de cada jugador.
 - 2.1.2. Se obtienen datos de una tabla de estadísticas.
 - 2.2. A continuación, se vuelve a la página gol.gg para recoger más datos de la partida.
 - 2.2.1. Se recogen los datos del sumario del mapa, como quien ha ganado, o los campeones jugados.
 - 2.2.2. Se accede a una tabla de estadísticas de cada jugador de la partida que da información más profunda del desempeño de los jugadores ese mapa.

A continuación, se presentan las variables del dataset, como aclaración, allí donde pone \$(color) hace referencia a que tenemos esa variable una vez por cada equipo (rojo o azul), pero que para cada equipo representa lo mismo, además, donde pone \$(posición), hace referencia a que

tenemos esa variable una vez por cada posición (top, junga, medio, adc o sup). Es importante aclarar que si una variable tiene \$(color) y \$(posición) en el nombre, tenemos esa variable 10 veces, una por cada posición por equipo, pero aquí solo aparece descrita una vez, porque siempre representa lo mismo. En concreto, los datos que se presentan en el dataset son los siguientes:

- torneo, el torneo al que pertenece el mapa.
- parte, la parte del torneo al que pertenece el mapa.
- fecha, el día en el que se jugó el mapa.
- semana, la semana o etapa de la parte del torneo en la que se juega el mapa.
- tiempo, la duración en minutos y segundos de la partida.
- parche, versión del juego en el que se jugó la partida.
- nombre_azul, nombre del equipo del lado azul.
- nombre_rojo, nombre del equipo del lado rojo.
- gana_azul, indica con un 1 si gana el equipo azul.
- gana_rojo, indica con un 1 si gana el equipo rojo.
- num_asesinatos_\$(color), el número de asesinatos que ha conseguido un equipo en la partida.
- primera_sangre, nombre del equipo que ha conseguido la primera sangre.
- num_torres_\$(color), número de torres destruidas por el equipo.
- primera_torre, equipo que ha conseguido destruir la primera torre.
- num_dragones_\$(color), número de dragones asesinados por equipo.
- num_dragones_viento_\$(color), número de dragones de viento asesinados por equipo.
- num_dragones_infierno_\$(color), número de dragones de infierno asesinados por equipo.
- num_dragones_oceano_\$(color), número de dragones de océano asesinados por equipo.
- num_dragones_montaña_\$(color), número de dragones de montaña asesinados por equipo.
- num_nashors_\$(color), número de nashors asesinados por equipo.
- num_oro_\$(color), cantidad de oro conseguida por el equipo a final de partida.
- ban_1_\$(color), primer campeón baneado de un equipo.
- ban_2_\$(color), segundo campeón baneado de un equipo.
- ban_3_\$(color), tercer campeón baneado de un equipo.
- ban_4_\$(color), cuarto campeón baneado de un equipo.
- ban_5_\$(color), quinto campeón baneado de un equipo.
- pick_1_\$(color), primer campeón seleccionado de un equipo. El del top.
- pick_2_\$(color), segundo campeón seleccionado de un equipo. El del jungla.
- pick_3_\$(color), tercer campeón seleccionado de un equipo. El del medio.
- pick_4_\$(color), cuarto campeón seleccionado de un equipo. El del adc.
- pick_5_\$(color), quinto campeón seleccionado de un equipo. El del sup.
- top_\$(color), el jugador de un equipo que juega en top.
- jng_\$(color), el jugador de un equipo que juega en jungla.
- mid_\$(color), el jugador de un equipo que juega en mid.
- adc_\$(color), el jugador de un equipo que juega en adc.
- sup_\$(color), el jugador de un equipo que juega en sup.

- kills_\$(posición)_\$(color), el número de asesinatos que ha obtenido un jugador de un equipo.
- assists_\$(posición)_\$(color), el número de asistencias que ha obtenido un jugador de un equipo.
- deaths_\$(posición)_\$(color), el número de muertes que ha obtenido un jugador de un equipo.
- summoner_1_\$(posición)_\$(color), el primer hechizo de invocador que tiene un jugador de un equipo.
- summoner_2_\$(posición)_\$(color), el segundo hechizo de invocador que tiene un jugador de un equipo.
- css_\$(posición)_\$(color), el número de súbditos asesinados por un jugador de un equipo.
- wards_destroyed_\$(color), el número de guardianes de visión eliminados por equipo.
- wards_placed_\$(color), el número de guardianes de visión colocados por equipo.
- jng_share_15_\$(color), el porcentaje de jungla asesinada por equipo en el minuto 15.
- jng_share_\$(color), el porcentaje de jungla asesinada por equipo al final de partida.
- diferencia_oro_\$(color), la diferencia de oro respecto al otro equipo cada 5 minutos de partida. Extraído a partir de la interacción con una gráfica de puntos.
- oro_\$(color), el oro que tiene un equipo cada 5 minutos de partida. Extraído a partir de la interacción con una gráfica de puntos.
- gold_\$(posición)_\$(color), el oro que tiene un jugador de un equipo cada 5 minutos de partida. Extraído a partir de la interacción con una gráfica de puntos.
- heraldos_\$(color), el número de heraldos que consiguió asesinar cada equipo.
- inhibs_\$(color), el número de inhibidores que destruyo cada equipo.
- First_Blood, indica que jugador de que equipo consiguió la primera sangre.
- Total_Damage_Dealt_\$(color)_\$(posicion), el daño total realizado por cada jugador de cada equipo.
- Physical_Damage_Dealt_\$(color)_\$(posicion), el daño físico realizado por cada jugador de cada equipo.
- Magic_Damage_Dealt_\$(color)_\$(posicion), el daño físico realizado por cada jugador de cada equipo.
- True_Damage_Dealt_\$(color)_\$(posicion), el daño verdadero realizado por cada jugador de cada equipo.
- Total_Damage_Objectives_\$(color)_\$(posicion), el daño verdadero realizado a objetivos por cada jugador de cada equipo.
- Damage_Taken_\$(color)_\$(posición), el daño recibido por cada jugador de cada equipo.
- Physical_Damage_Taken_\$(color)_\$(posición), el daño físico recibido por cada jugador de cada equipo.
- Magic_Damage_Taken_\$(color)_\$(posición), el daño mágico recibido por cada jugador de cada equipo.
- True_Damage_Taken_\$(color)_\$(posición), el daño verdadero recibido por cada jugador de cada equipo.
- cs_in_jung_team_\$(color)_\$(posicion), el número de monstruos de la jungla de su equipo asesinados por cada jugador de cada equipo.
- cs_in_jung_enemy_\$(color)_\$(posicion), el número de monstruos de la jungla del equipo enemigo asesinados por cada jugador de cada equipo.

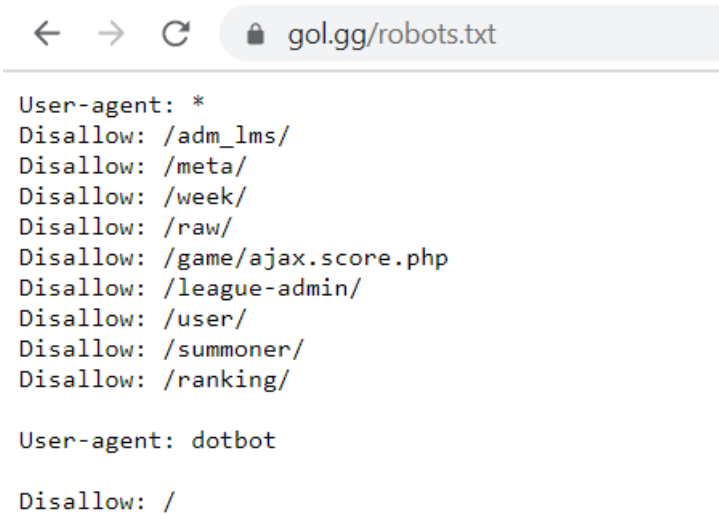
- CSM_\$(color)_\$(posición), el número de súbditos por minuto asesinados por cada jugador de cada equipo.
- Golds_\$(color)_\$(posición), el oro final por cada jugador de cada equipo.
- GPM_\$(color)_\$(posición), el número de súbditos por minuto asesinados por cada jugador de cada equipo.
- GOLD_\$(color)_\$(posición), el porcentaje de oro total por cada jugador de cada equipo respecto al oro de su equipo.
- Vision_Score_\$(color)_\$(posición), la puntuación de visión de cada jugador de cada equipo.
- Wards_placed_\$(color)_\$(posición), el número de guardianes de visión puestos por cada jugador de cada equipo.
- Wards_destroyed_\$(color)_\$(posición), el número de guardianes de visión destruidos por cada jugador de cada equipo.
- Control_Wards_\$(color)_\$(posición), el número de guardianes de control de visión puestos por cada jugador de cada equipo.
- VS%_\$(color)_\$(posición), el porcentaje de puntuación de visión total por cada jugador de cada equipo respecto a la puntuación de visión de su equipo.
- Total_damage_Champios_\$(color)_\$(posición), el total de daño realizado a enemigos por cada jugador de cada equipo.
- Physical_damage_Champios_\$(color)_\$(posición), el total de daño físico realizado a enemigos por cada jugador de cada equipo.
- Magic_damage_Champios_\$(color)_\$(posición), el total de daño mágico realizado a enemigos por cada jugador de cada equipo.
- True_damage_Champios_\$(color)_\$(posición), el total de daño verdadero realizado a enemigos por cada jugador de cada equipo.
- DPM_\$(color)_\$(posición), el daño por minuto por cada jugador de cada equipo.
- DMG_\$(color)_\$(posición), el porcentaje de daño total por cada jugador de cada equipo respecto al daño total de su equipo.
- Solo_kills_\$(color)_\$(posición), el número de asesinatos en solitario realizados por cada jugador de cada equipo.
- Double_kills_\$(color)_\$(posición), el número de asesinatos dobles realizados por cada jugador de cada equipo.
- Triple_kills_\$(color)_\$(posición), el número de asesinatos triples realizados por cada jugador de cada equipo.
- Quadra_kills_\$(color)_\$(posición), el número de asesinatos cuádruples realizados por cada jugador de cada equipo.
- Penta_kills_\$(color)_\$(posición), el número de asesinatos quíntuples realizados por cada jugador de cada equipo.
- CSD@15_\$(color)_\$(posición), la diferencia en súbditos asesinados respecto a su enemigo en su posición por cada jugador de cada equipo.
- XPD@15_\$(color)_\$(posición), la diferencia en experiencia respecto a su enemigo en su posición por cada jugador de cada equipo.
- LVLD@15_\$(color)_\$(posición), la diferencia en niveles respecto a su enemigo en su posición por cada jugador de cada equipo.
- Damage_towers_\$(color)_\$(posición), el daño a torres realizado por cada jugador en cada partida.

- heal_\$(color)_\$(posición), la cantidad de curación realizada por cada jugador en cada partida.
- ccing\$(color)_\$(posición), el tiempo de paralizaciones, ralentizaciones, etc realizado por cada jugador en cada partida.

Agradecimientos.

Los datos obtenidos pertenecen a gol.gg y a leagueoflegends.com. La principal web utilizada, gol.gg, es una página dedicada a almacenar las estadísticas de las partidas de jugadores profesionales de League of Legends, con el objetivo de tener organizados los datos por torneos, equipos y jugadores y poder identificar las estadísticas más llamativas. Por otra parte, se accede al historial de partida de leagueoflegends.com, que es la fuente oficial de los datos de la partida que proporciona la empresa desarrolladora del video juego (Riot Games), y por tanto es la fuente oficial de los datos.

Para poder obtener los datos, primero he estudiado los ficheros robots.txt. Respecto a el fichero de gol.gg, tenemos:



```
User-agent: *
Disallow: /adm_lms/
Disallow: /meta/
Disallow: /week/
Disallow: /raw/
Disallow: /game/ajax.score.php
Disallow: /league-admin/
Disallow: /user/
Disallow: /summoner/
Disallow: /ranking/

User-agent: dotbot

Disallow: /
```

Vemos como para los usuarios normales, o cualquiera que no sea dotbot, se puede acceder a las páginas de torneos o partidas, que son <https://gol.gg/tournament/> y <https://gol.gg/game>.

Para leagueoflegends.com no he conseguido encontrar un robots.txt, pero si lo he conseguido para la versión coreana de la web, por lo que entiendo que se puede hacer web scraping a cualquier versión, ya que el historial de partida es el mismo y los datos son los mismos.

← → ↻ 🔒 leagueoflegends.co.kr/robots.txt

```
User-agent: *
Allow: /
Allow: /?m=esports
Allow: /?m=news
Allow: /?m=forum
Allow: /?m=contribute
Allow: /?m=rules
Allow: /?m=multimedia
Allow: /?m=riot-store
Disallow: /?m=
Disallow: /*schwrld=
Disallow: /*keyword=

User-agent: Twitterbot
Disallow:
```

Vemos como se especifica que cualquier usuario puede acceder a las webs de leagueoflegends. Salvo aquellas que están en Disallow, y el bot de Twitterbot. Nosotros queremos acceder a matchhistory, que está permitido.

Es importante comentar que Riot Games especifica que con su Propiedad Intelectual permite desarrollar proyectos de fan gratuitos en beneficio de la comunidad, y establece una serie de reglas en: <https://www.riotgames.com/es/legal>.

Inspiración.

El conjunto de datos obtenido se corresponde a una gran cantidad de datos de partidas profesionales de League of Legends. El conjunto de datos es muy interesante sobre todo en el área de minería de datos y en la modelización predictiva, ya que con el dataset es posible plantearse responder preguntas de todo tipo:

- ¿Cuál es el mejor jugador en una posición? Ya que podemos estudiar muchas variables de desempeño de un jugador en la partida, podemos identificar si destaca el jugador frente a otros, incluso en las derrotas de su equipo. Esto es muy importante para fichar jugadores por parte de equipos, que podrían estudiar mediante datos a qué jugador fichar.
- ¿Cómo consiguen ganar los equipos y que debilidades tienen? Al tener variables de los equipos y de sus jugadores, podemos realizar un perfilado de equipo, identificando aquellos puntos clave de un equipo, por ejemplo, jugar a obtener dragones y objetivos neutrales y no realizando peleas de equipo que puedan ser beneficiosas para el rival.
- ¿Qué equipo ganaría entre dos equipos? Al tener variables de los equipos y jugadores, podemos estudiar el perfil de cada equipo y generar un modelo predictivo que prediga que equipo va a ganar en función de la trayectoria de cada equipo, sus jugadores estrella, y sus enfrentamientos directos.

Un aspecto interesante del conjunto de datos, es que tiene una gran cantidad de variables por partida, esto permite que se pueda responder a las preguntas anteriormente propuestas y a muchas más al procesar los datos en función del objetivo que se quiera resolver.

El conjunto de datos es muy interesante también, porque actualmente no hay un sitio abierto donde los usuarios puedan coger datos de las partidas profesionales de League of Legends,

mientras que si hay una API abierta para partidas no profesionales. Esto permite también identificar las diferencias de juego entre partidas profesionales y partidas no profesionales.

Licencia.

Debido a que Riot Games es en definitiva el propietario de la propiedad intelectual en base a la que se ha generado este dataset, y que en <https://www.riotgames.com/es/legal> establece que con su propiedad intelectual está permitido desarrollar proyectos de fan gratuitos en beneficio de la comunidad. Lanzo este dataset bajo la licencia: **Released Under CC0: Public Domain License.**

Abriendo así el dataset a dominio público y renunciando a mis derechos a la obra. Ya que considero que es el deseo de Riot Games que los proyectos en torno a su propiedad intelectual sean abiertos.

Cualquiera puede copiar, modificar, distribuir e interpretar este dataset, sin pedir permiso a mi persona. Eso sí, esto no afecta en ninguna forma los derechos de patentes o de marcas sobre la obra, ni derechos que otras personas puedan tener en la obra o en cómo la obra es usada, por tanto, las limitaciones sobre aquello que se puede o no hacer con el dataset vienen controladas por Riot Games en <https://www.riotgames.com/es/legal> y antes de realizar cualquier obra con el dataset es importante consultar la página web.

Código.

El código utilizado para generar el dataset, se encuentra en el repositorio de github, concretamente en el directorio code:

https://github.com/manurubo/LeagueOfLegends_ProfessionalGames

La explicación de los ficheros del código y las aclaraciones sobre aquellas partes más difíciles del código y como las he solucionado se encuentran en el fichero readme del repositorio github.

Dataset.

El dataset tiene el DOI: 10.5281/zenodo.4265268

Se encuentra en la web:

<https://zenodo.org/record/4265268#.X6mwUmhKhPY>

Contribuciones.

Manuel Ruiz Botella -> MRB

Contribuciones	Firma
Investigación previa	MRB
Redacción de las respuestas	MRB
Desarrollo código	MRB