

Modelo de predicciones Counter-Strike

Análisis de desempeño de los equipos e impacto en las apuestas

Integrantes: Vicente Araya, Aldo Coello, Sebastián Jiménez, Carolina Neira

Curso: Introducción a la Ciencia de Datos (IMT2200-1)

Contexto y motivación

Counter-Strike 2 o CS2 es un juego multijugador de disparos en primera persona (o FPS por sus siglas en inglés) en el cual dos equipos de cinco jugadores cada uno, se enfrentan en una partida de 24 rondas en alguno de los 7 mapas activos, donde el primero en ganar 13 de ellas se lleva la victoria. Las partidas competitivas están divididas en dos mitades de 12 rondas cada una, en la primera mitad, uno de los equipos juega como el bando antiterrorista y el otro como el bando terrorista, y en la siguiente mitad, los equipos cambian de bandos. En cada ronda los terroristas tienen como objetivo plantar una bomba en alguno de los dos puntos especiales del mapa o bien eliminarlos a todos los antiterroristas. Por otro lado, los antiterroristas tienen como objetivo defender estos puntos por la duración de la ronda, desactivar la bomba en caso de ser plantada o bien eliminar a todos los terroristas.

El balance de toda partida está dado por la economía que mantienen los equipos a lo largo de esta, comenzado con 800 dólares al inicio de esta y ganando más según cómo se desarrollan las rondas. Con este dinero, cada bando puede comprar armas de diferentes tipos y utilidades tales como granadas, humos, armadura o kit de desactivación con el fin de obtener una ventaja sobre el equipo rival, sin embargo, si un jugador muere con su equipamiento, tiene que volver a comprarlo a la siguiente ronda, lo que no siempre es posible.

Una de las mayores diferencias que la saga Counter-Strike posee por sobre otros FPS, son sus mecánicas de disparo. En CS2 cada arma posee un patrón de dispersión fijo y único, por lo que no importa quien dispare el arma, el resultado siempre es el mismo. Ahora bien, para que esto sea verdad, el jugador tiene que estar quieto, el disparar mientras se está en movimiento perjudica su precisión, volviendo su patrón de disparo impredecible. Estas dos mecánicas, junto a lo dicho anteriormente, propician las condiciones perfectas para crear uno de los juegos competitivos más grandes de la historia, con una gran escena profesional que ofrece premios de miles de dólares.

De entre todos los torneos profesionales de CS2 realizados cada año, el más importante es el Major, donde 24 de los mejores equipos del mundo compiten por premios de hasta medio millón de dólares. En un formato de clasificatorias de grupos según nivel y posteriormente en eliminatorias.

Los partidos del Major son los mejores momentos para apostar debido al gran volumen de personas interesadas, siendo estas mismas impulsadas por los mismos desarrolladores del juego. Esto es gracias a la publicidad del evento dentro del mismo juego y a la posibilidad de participar en el “Pick'em”, un pequeño juego en el que por acertar a los equipos clasificados y los resultados de las eliminatorias se pueden obtener recompensas que venden a precios elevados en los mercados de la comunidad. Por estas razones, existen sitios web que recopilan las estadísticas de los partidos, equipos y jugadores profesionales, la más conocida, HLTV.

Es debido a la multitudinaria audiencia y su afición al juego, que nos interesa estudiar en profundidad el funcionamiento de las apuestas y cómo éstas se ven influenciadas por las distintas aristas que conforman el escenario de una partida, para retornar finalmente la combinación “ganadora” (si hay una que exista, y cuán probable es que suceda). Nos parece un gran desafío, pero a la vez, una gran y divertida experiencia.

Objetivos

- Desarrollar un modelo predictivo que pueda predecir los resultados de partidos profesionales de Counter-Strike 2.
- Identificar las variables más influyentes en el desempeño de los equipos (economía, selección de mapas, lugar del torneo, etc.)
- Predecir el campeón del Major.
- Proporcionar una visión a los apostadores y jugadores sobre los factores que influyen en el resultado de un partido para mejorar las decisiones.

Datos

Para este proyecto consideraremos los datos proporcionados por las siguientes dos páginas:

- Página principal: <https://www.hltv.org/stats>
- Página secundaria: <https://liquipedia.net/counterstrike/Portal:Statistics>
- Repositorio donde estarán los datos: [Jimenez73/Proyecto-Semestral-Grupo-13: Proyecto semestral del grupo 13 para el ramo Intro a la Cs de Datos \(github.com\)](https://github.com/Jimenez73/Proyecto-Semestral-Grupo-13)

Todos los datos de los equipos, jugadores, partidos y torneos están organizados en tablas a través de diversas pestañas de la página web de HLTV. Además, consideramos la página Liquipedia en caso de datos faltantes.

Las siguientes variables serán las que pensamos que usaremos para el proyecto:

1. Desempeño económico de los equipos durante la partida
2. Mapas seleccionados

3. Armas utilizadas
4. Aplicación y eficiencia de utilidad
5. Estadísticas de rondas ganadas
6. Estadísticas de los jugadores

Formato preliminar: CSV obtenido mediante web scraping.

Volumen: Un año de datos de los partidos profesionales.

Preguntas de investigación

- Variables que afectan en el juego:
 - ¿Qué factores dentro del juego tienen mayor impacto en el resultado final?
 - ¿Cómo influye el desempeño económico de cada equipo durante la partida para predicción del resultado final?
- Considerando selección de mapas:
 - ¿Hay cierta correlación entre los mapas elegidos y el porcentaje de victorias de cada equipo?
- Ubicación del torneo y presencia de público:
 - ¿Hay diferencia significativa en el rendimiento de los equipos que juegan torneos virtuales y aquellos que lo hacen presencial?
 - ¿De qué manera influye la presencia del público en vivo el desempeño de los jugadores?

Diseño tentativo

- Recolección de datos mediante formato descrito.
- Limpieza y procesamiento de datos: Se realizará una limpieza de los datos recolectados, estructurándolos en DataFrame de Pandas para facilitar el análisis.
- Análisis exploratorio: Explorar cada uno de los datos para entender las relaciones entre ellos mediante descripciones estadísticas.
- Uso de modelos predictivos.
- Evaluación y verificación del modelo.
- Visualización de los datos (Uso de Matplotlib y Seaborn).