

**UNIVERSIDAD DEL VALLE DE GUATEMALA**

Data Science

Sección 10



# **Proyecto 1**

## **Análisis Exploratorio**

JESSICA PAMELA ORTIZ IXCOT 20192

ESTEBAN ALDANA GUERRA 20591

JUAN CARLOS BAJAN CASTRO 20109

JOSÉ RODRIGO BARRERA GARCÍA 20807

**GUATEMALA, Septiembre 2023**

## **Situación Problemática:**

Desde temprana edad, muchos jóvenes con talento dedican tiempo, dinero y esfuerzo en entrenar para llegar a ser profesionales en el fútbol. A pesar de su dedicación, muchos de estos jóvenes comienzan su carrera en ligas juveniles o semi-profesionales, las cuales frecuentemente carecen de recursos adecuados. Uno de los principales recursos que falta es la capacidad de recoger datos de eventos durante los partidos. Estos datos son esenciales para generar análisis detallados sobre el rendimiento de los equipos y jugadores, lo cual puede influir en las decisiones sobre su carrera futura.

En su estado actual, la recolección de estos datos se realiza de forma manual, lo que implica un proceso laborioso y con alta probabilidad de errores. Requiere múltiples pasos y la intervención de numeroso personal para recoger, verificar y analizar los datos. Debido a la inversión en tiempo y recursos que este proceso manual demanda, generalmente solo se lleva a cabo en competencias de nivel profesional, dejando a las ligas menores sin acceso a esta herramienta valiosa.

Esto presenta un problema, ya que sin la recolección precisa y eficiente de datos, jugadores talentosos en estas ligas menores podrían ser pasados por alto y no recibir las oportunidades que merecen. Además, las ligas y entrenadores no cuentan con la información necesaria para tomar decisiones informadas y mejorar sus tácticas y estrategias.

Por lo tanto, existe una necesidad evidente de innovar y encontrar una solución más eficiente y efectiva para recopilar y analizar datos de eventos en el fútbol, especialmente en las ligas que no tienen los medios para hacerlo manualmente.

## **Problema científico**

¿Cómo se puede desarrollar un método automático y eficiente para la detección y recopilación de datos de eventos en partidos de fútbol de ligas juveniles y semi-profesionales, que tradicionalmente carecen de los recursos para llevar a cabo esta tarea manualmente, garantizando precisión, rapidez y profundidad en el análisis, para asegurar que ningún jugador talentoso sea pasado por alto y que los equipos y entrenadores tengan acceso a información valiosa para mejorar su rendimiento?

## **Objetivos:**

### **Objetivo General:**

- Desarrollar un sistema automatizado para la detección y recopilación de datos de eventos en partidos de fútbol en ligas juveniles y semi-profesionales que mejore la precisión y eficiencia del proceso actual.

## Objetivos Específicos:

- Diseñar un algoritmo basado en inteligencia artificial que pueda identificar y registrar eventos claves del partido (como jugadas, entadas, saques de banda, entre otros) con una precisión del 80% o más.
- Implementar una herramienta que, utilizando el algoritmo diseñado, pueda procesar y analizar los datos recolectados en tiempo real o post-partido, ofreciendo reportes detallados y fáciles de interpretar para entrenadores y analistas deportivos.

## Descripción de los datos

### 1. video\_id:

**Descripción:** Es una identificación única asignada a cada video dentro de las carpetas 'train' y 'test'.

**Tipo:** Alfanumérico.

**Uso:** Permite correlacionar los eventos y sus atributos en el excel con los clips de vídeo específicos.

### 2. time:

**Descripción:** Representa un punto temporal en el video, medido en segundos. Es el momento exacto en el cual ocurre un evento o comienza/termina un clip que es de interés para el análisis.

**Tipo:** Numérico (decimal).

**Uso:** Ayuda a ubicar con precisión el instante en el video donde ocurre un evento en particular.

### 3. event:

**Descripción:** Es una etiqueta que define el tipo de actividad o acción que está teniendo lugar en ese momento específico del video.

**start:** Indica el comienzo de un clip o segmento de interés.

**end:** Señala el final de ese segmento de interés.

**play:** Un evento donde el balón está en juego.

**throwin:** Representa un saque de banda.

**challenge:** Hace referencia a un desafío o enfrentamiento entre jugadores, por ejemplo, al disputar un balón.

**Tipo:** Categórico.

**Uso:** Permite categorizar y segmentar los momentos clave dentro del video.

#### 4. event\_attributes:

**Descripción:** Es una lista que contiene características adicionales o especificaciones relacionadas con el evento en cuestión. Estas etiquetas ofrecen un nivel más detallado de análisis sobre el tipo y naturaleza del evento.

**Ejemplos:**

**opponent\_dispossessed:** El oponente fue despojado del balón.

**pass:** Un pase entre jugadores.

**openplay:** Una jugada en curso normal, sin interrupciones.

**possession\_retained:** El equipo mantiene la posesión del balón.

**ball\_action\_forced:** Se obligó a una acción con el balón.

**challenge\_during\_ball\_transfer:** Un desafío o enfrentamiento durante una transferencia o pase de balón.

**cross:** Un centro al área.

**corner:** Un saque de esquina.

**freekick:** Un tiro libre.

**fouled:** Se cometió una falta sobre el jugador.

**opponent\_rounded:** El oponente fue superado o driblado.

**Tipo:** Lista de valores categóricos.

**Uso:** Ayuda a comprender y analizar con mayor profundidad el contexto y detalles de un evento en particular.

Para acceder a los análisis y procesos de limpieza de datos, consulta el Jupyter Notebook correspondiente. Este documento contiene un registro detallado de todas las operaciones realizadas en relación con los datos. A través del Jupyter Notebook, podrás examinar de cerca las etapas de limpieza y análisis que se han llevado a cabo.

### Hallazgos y conclusiones

En el marco de nuestro análisis exploratorio, hemos identificado diversas características clave. En primer lugar, hemos destacado tres categorías principales de eventos observables: saques de banda, jugadas y entradas. Estas representan las principales facetas de nuestra base de datos.

Posteriormente, hemos desglosado esta información en múltiples clips o conjuntos de datos de entrenamiento, categorizados según los diferentes tipos de jugadas. Este enfoque nos permitirá llevar a cabo un análisis en profundidad. Además, hemos examinado la cantidad de eventos registrados en cada partido.

Adicionalmente, hemos explorado la relación temporal entre eventos dentro de los videos y hemos evaluado las interacciones entre variables. Para obtener un análisis detallado de estos resultados, le invitamos a consultar el Jupyter Notebook adjunto.

En la próxima fase del proyecto, aplicaremos diversas técnicas para identificar las distintas etapas del juego y determinar qué información valiosa podemos extraer de ellas. Emplearemos enfoques como las redes neuronales convolucionales (CNN) para analizar las imágenes, así como herramientas como YOLO para realizar un análisis más exhaustivo de los diferentes videos.