

1. Objetivo General

Realizar un análisis de los partidos de futbol entre dos países para poder predecir quién podría ser el ganador en su próximo encuentro.

2. Propuesta y Organización CRISP-DM

2.1. Business Understanding (Entendimiento del negocio/problema)

- Define claramente cuál es el problema de negocio o de investigación.

Se necesita saber el posible ganador del próximo partido de Ecuador vs Brasil.

- Describe el contexto: ¿cuál es la hipótesis principal?, ¿qué necesitas comprobar o resolver?, ¿cuál es tu pregunta de investigación?

Tengo una base con los resultados de todos los partidos de futbol desde 1872 hasta 2025

En base a esta información se requiere predecir el próximo ganador en el partido que se va a llevar a cabo entre Ecuador y Brasil en junio de 2025

Mi pregunta sería quién será el ganador del siguiente partido entre Ecuador y Brasil

- Explica el valor que proporcionará tu producto de datos MVP (¿por qué es relevante para tu institución, empresa o investigación?).

Es importante para saber si vamos a poder ir al mundial 2026

2.2. Data Understanding (Entendimiento de los datos)

- Explica de dónde provienen tus datos (fuentes, tipo de datos, frecuencia de actualización, etc.).

La información proviene de la plataforma Kaggle www.kaggle.com

Esta información se actualiza semanal o diariamente depende de los partidos llevados a cabo.

La base utilizada se la bajó un día después del partido Ecuador vs Brasil, el 26 de marzo de 2025, y ya tenía la información actualizada del 25 de marzo de 2025.

El archivo es .csv

Los tipos de datos son:

```
dtype: object
date          datetime64[ns]
home_team      object
away_team      object
home_score     int64
away_score     int64
tournament     object
city           object
country        object
neutral        bool
```

- Describe de manera general las variables y su posible relevancia o relación con el problema.

date: es la fecha en la que se llevó a cabo el partido

home_team: es el equipo que juega en casa

away_team: es el equipo que juega de visitante

home_score: es el puntaje del equipo que juega en casa

away_score: es el puntaje del equipo que juega de visitante

tournament: el torneo en el que participaron

city: la ciudad donde se llevó a cabo el encuentro

country: el país donde se llevó a cabo el encuentro

neutral: indica si el partido se llevó a cabo en un lugar neutral

De los campos indicados home_team, away_team, home_score y away_score son claves para el análisis

- Identifica los posibles desafíos: datos faltantes, duplicados, inconsistencias, calidad y confiabilidad.

Los datos son confiables y sin inconsistencias, se validaron los datos de los partidos para verificar su veracidad. Tampoco tiene duplicidad

2.3. Data Preparation (Preparación de los datos)

- Lista las tareas de limpieza y transformación necesarias (Data Wrangling).

Se debe revisar el tipo de dato de los campos

Se debe verificar que no existan duplicados con la función duplicated

Se debe contar las filas y columnas

Se debe validar valores faltantes con la función isnull

Se debe asegurar de que la fecha sea tipo fecha con la función to_datetime

Se debe ordenar por fecha con la función sort_values

Se debe crear promedios de puntajes con la función mean

- Documenta la extracción de datos y la manipulación para llegar a la forma deseada.

2.4. Modeling (Modelado)

2.5. Evaluation (Evaluación)

2.6. Deployment (Despliegue)

