



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Países Clasificados al Mundial Qatar 2022 Predicción

Proyecto Cs Datos IMT2200



Profesor:

Paula Aguirre Aparicio

Estudiantes:

Vicente Espinoza Flores

Rubén Marambio Miranda

Germán Vilos Corral

Sobre el contexto y motivación:

Es bien sabido, que el fútbol es el principal deporte en el mundo actual. Así lo reflejan sin fin de estudios que concluyen el hecho de que eventos como el mundial de selecciones, paraliza el mundo. En concreto, para nadie es indiferente este deporte. Ante esta realidad, se han forjado un sinfín de industrias que se sustentan en torno al fútbol, entiéndase por esto, los paneles de periodistas dedicados exclusivamente a este deporte, las aplicaciones móviles de apuesta, entre otras.

A partir de esto, es entendible el inmenso valor que adquiere la capacidad de predecir resultados. Dicha tarea no es trivial, en efecto, para un partido hay más de 50 personas con participación (tanto jugadores en banca, como utileros) activa, sumando a las condiciones ambientales, como otras. Pese a esto, pareciera haber selecciones con una mayor probabilidad de éxito, a pesar de poseer jugadores con regímenes similares, los cuales pueden hasta participar de las mismas ligas. En efecto, el rendimiento de una selección de fútbol puede ser entendido a partir de su rendimiento histórico.

En este trabajo, utilizaremos un conjunto de datos sobre resultados futbolísticos para intentar predecir el rendimiento actual de las selecciones tomando en cuenta su rendimiento en partidos de visitante, partidos de local y participaciones en instancias finales de copas del mundo. A partir de estos valores, creamos un modelo que genera índices por cada equipo, que utilizaremos para intentar entender el rendimiento actual de los equipos en las clasificatorias a Qatar 2021.

Objetivos:

- Entender el rendimiento histórico de un equipo a partir de sus resultados.
- Entender el rendimiento actual de una selección de fútbol en el proceso clasificatorio al mundial de 2022.
- Aplicar este estudio y comprender patrones de rendimiento que puedan darse en ciertos lugares.
- Aplicar nuestros conocimientos en Ciencia de datos para los procesos de extracción, manejo y visualización de datos.

Diseño tentativo:

Utilizando el Dataset: “International football results from 1872 to 2017”, planeamos calcular la cantidad de victorias de local que tiene cada equipo por el total de partidos de local, la cantidad de victorias como visitante por partidos de visitante. Luego, con el segundo Dataset, calcularemos la cantidad de victorias de cada equipo en instancias de finales por la cantidad de finales disputadas.

Utilizaremos estos indicadores para calcular el siguiente índice a cada equipo:

$$I = \frac{victorias_{local}}{partidos_{local}} * 0.3 + \frac{victorias_{visita}}{partidos_{visita}} * 0.4 + \frac{victorias_{finales}}{partidos_{finales}} * 0.3$$

Luego, analizaremos dicho parámetro teniendo en consideración la posición actual de cada equipo en su respectiva tabla de clasificatoria.

Datos:

Muestra de partidos Mundial de futbol:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	Win conditions	Attendance	Half-time Home Goals	Half-time Away Goals	Referee	Assistant 1	Assistant 2	RoundID	MatchID	Home Team Initials	Away Team Initials
0	1930.0	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4.0	1.0	Mexico		4444.0	3.0	0.0	LOMBARDI Domingo (URU)	CRISTOPHE Henry (BEL)	REGO Gilberto (BRA)	201.0	1096.0	FRA	MEX
1	1930.0	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3.0	0.0	Belgium		18346.0	2.0	0.0	MACIAS Jose (ARG)	MATEUCCI Francisco (URU)	WARNKEN Alberto (CHI)	201.0	1090.0	USA	BEL
2	1930.0	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2.0	1.0	Brazil		24059.0	2.0	0.0	TEJADA Anibal (URU)	VALLARINO Ricardo (URU)	BALWAY Thomas (FRA)	201.0	1093.0	YUG	BRA
3	1930.0	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3.0	1.0	Peru		2549.0	1.0	0.0	WARNKEN Alberto (CHI)	LANGENUS Jean (BEL)	MATEUCCI Francisco (URU)	201.0	1098.0	ROU	PER
4	1930.0	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1.0	0.0	France		23409.0	0.0	0.0	REGO Gilberto (BRA)	SAUCEDO Ulises (BOL)	RADULESCU Constantin (ROU)	201.0	1085.0	ARG	FRA

Muestra Partidos 1872 – 2017:

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False