A decorative network diagram in the top right corner consisting of three blue circular nodes of varying sizes connected by thin grey lines. Ellipses (...) are placed at the ends of the lines, suggesting a larger network.

PREDICCIÓN DE RESULTADOS EN PARTIDOS DE FÚTBOL (PARL-AI)

Juan Pablo Pérez Leal
Laura Natalia Jaimes Antolínez

Tabla de Contenidos

01

Enfoque

Copa América 2021 .

02

Dataset

Características generales.

03

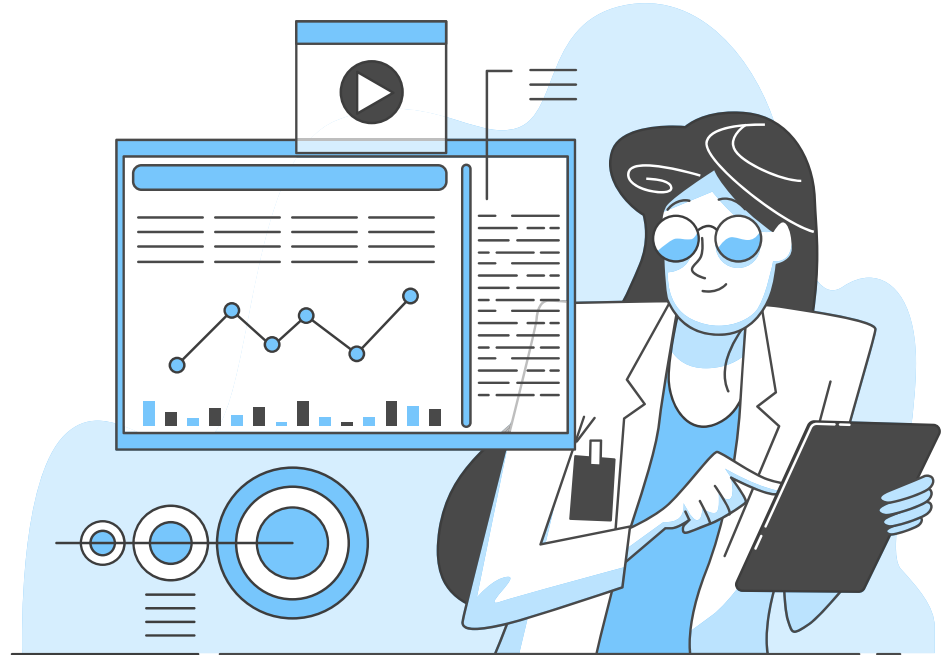
Metodología

Procesamiento de datos.
Modelamiento del problema.

04

Conclusiones

Resultados
Dificultades en el proyecto.



Enfoque

El objetivo de este proyecto se centra en comprobar los resultados de la Copa América 2021 usando diversos estimadores de clasificación y regresión.



Debido a las limitadas características de nuestro dataset, se espera que la precisión de nuestros resultados se va afectada



Dataset



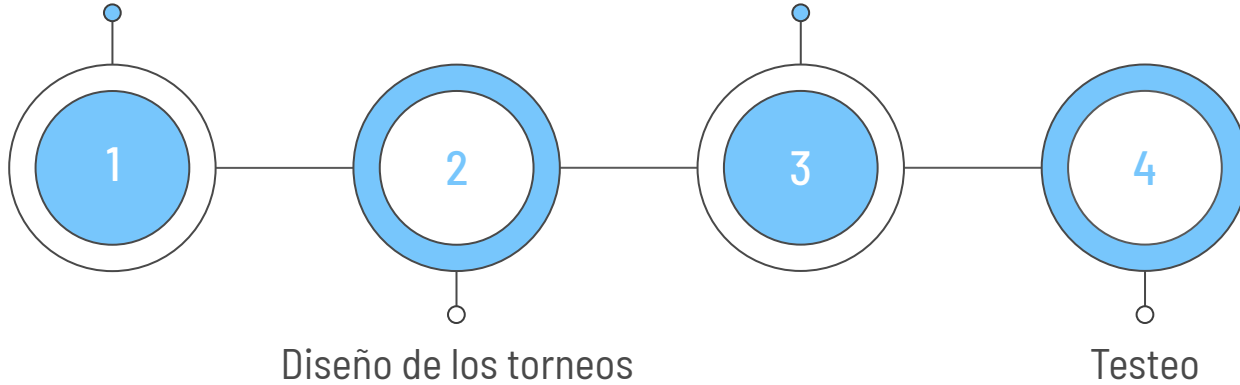
International Football Results from 1872 to 2021.

date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
09/09/2021	Colombia	Chile	3	1	FIFA World Cup qualification	Barranquilla	Colombia	FALSE
09/09/2021	Paraguay	Venezuela	2	1	FIFA World Cup qualification	Asunción	Paraguay	FALSE
09/09/2021	Uruguay	Ecuador	1	0	FIFA World Cup qualification	Montevideo	Uruguay	FALSE
09/09/2021	Argentina	Bolivia	3	0	FIFA World Cup qualification	Buenos Aires	Argentina	FALSE
09/09/2021	Brazil	Peru	2	0	FIFA World Cup qualification	Recife	Brazil	FALSE

Metodología

Procesamiento de los
datos

- Entrenamiento usando
- SVR (Support Vector Regressor)
 - SVC (Support Vector Classifier)
 - Random Forest



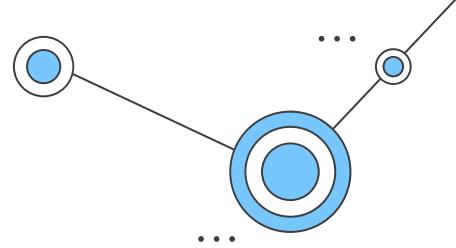


01

Procesamiento de datos

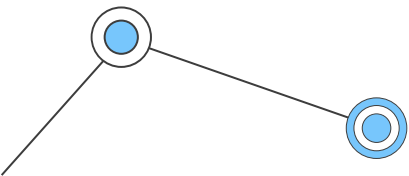


Extracción y Filtración



	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral	result
32058	2010-03-31	4	10	0	0	Friendly	Temuco	4	0	0
32239	2010-06-28	3	4	3	0	FIFA World Cup	Johannesburg	21	1	1
32267	2010-08-11	2	5	1	1	Friendly	La Paz	2	0	0
32348	2010-09-03	10	5	0	2	Friendly	Puerto la Cruz	10	0	2
32414	2010-09-07	10	6	1	0	Friendly	Barquisimeto	10	0	1


1. De los países latinoamericanos, se extraen todos los partidos que se disputaron desde el 2010 hasta la fecha.
2. Colocamos una etiqueta referente a cada país para facilitar la evaluación del modelo de inferencia y su respectivo resultado.





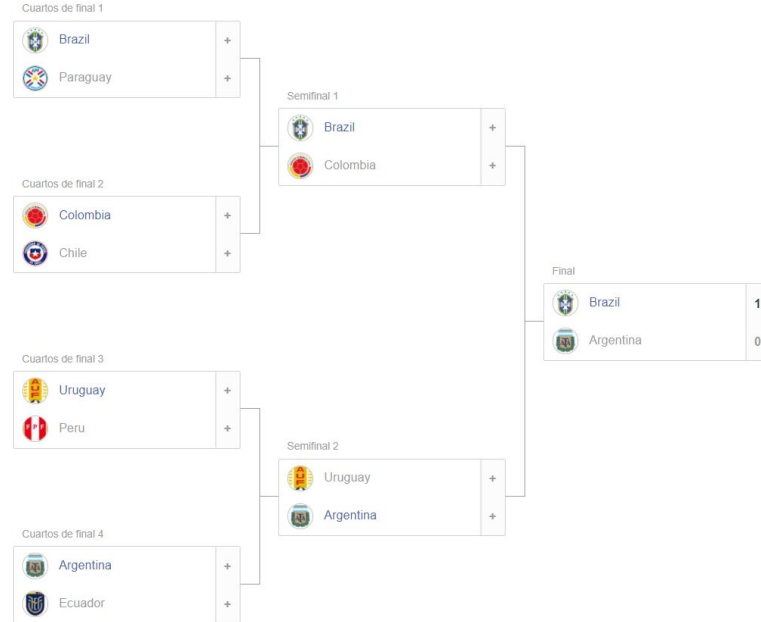
02

Diseño de Torneos

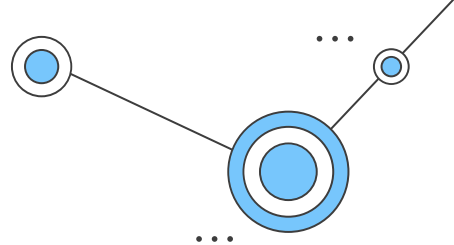


Copa América 2021











Para rectificar los resultados de la Copa América 2021, se seleccionan los equipos que jugaron este torneo y de acuerdo a los grupos preseleccionados se modela cada encuentro.

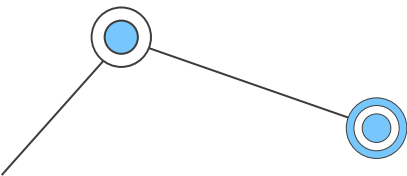


Eliminatorias Qatar 2022



De acuerdo a los partidos predefinidos para este torneo, se diseña el parámetro de "Resultado" en el que evaluaremos el ganador de cada encuentro.

Club	PJ	G	E	P	GF	GC	DG	Pts	Últimos 5
1  Brasil	4	4	0	0	12	2	10	12	✓✓✓✓✓
2  Argentina	4	3	1	0	6	2	-4	10	✓✓✓○
3  Ecuador	4	3	0	1	13	6	7	9	✓✓✓✓×
4  Paraguay	4	1	3	0	6	5	1	6	✓-✓-○
5  Uruguay	4	2	0	2	7	7	0	6	✓✓✓✓○
6  Chile	4	1	1	2	6	6	0	4	✓✓-✓×
7  Colombia	4	1	1	2	6	11	-5	4	✓✓-✓○
8  Venezuela	4	1	0	3	2	6	-4	3	✓✓✓×
9  Perú	4	0	1	3	4	10	-6	1	✓✓-○
10  Bolivia	4	0	1	3	5	12	-7	1	✓-✓×

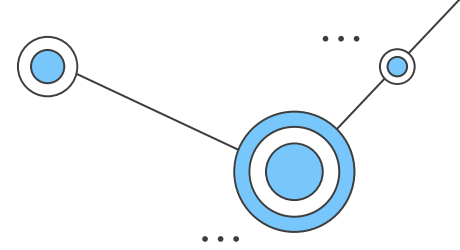


03

Entrenamiento

- Support Vector Machine (SVM)
- Support Vector Classifier (SVC)
- Random Forest

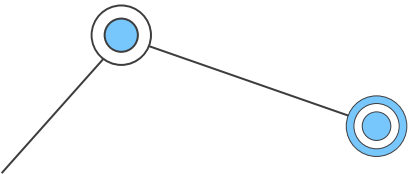
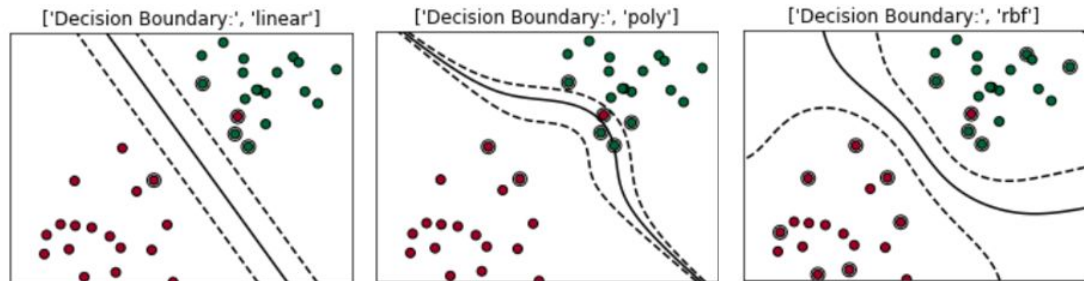
Support Vector Machine (SVM)



Para dar solución a nuestro problema de inferencia, el Support Vector Machine modela diferentes tipos de regresiones (Linear - Poly - RBF) con el fin de encontrar la forma más óptima de clasificar varias clases.

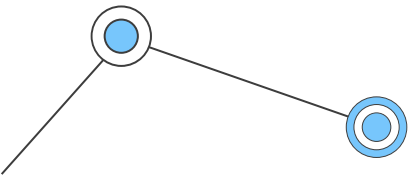
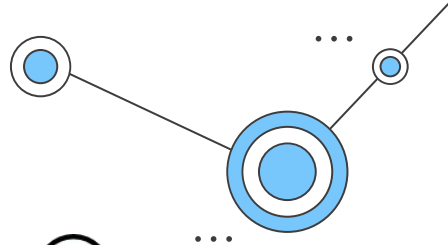
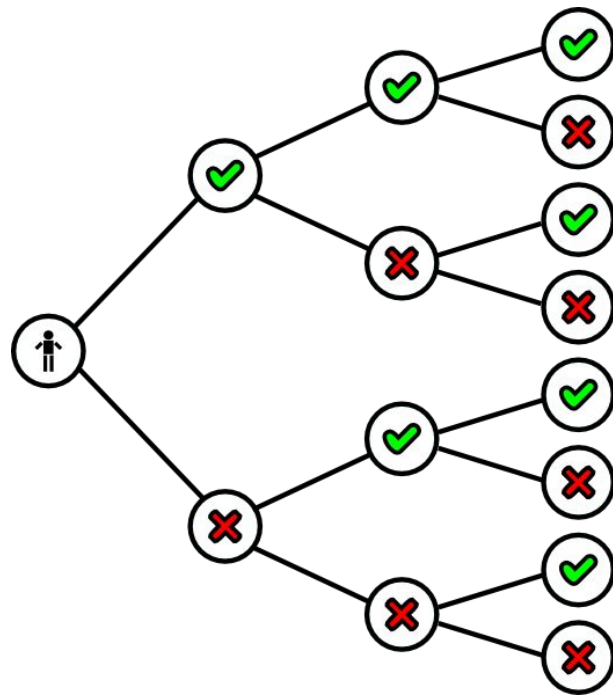
Para este primer método generamos el modelo y según el **Puntaje** de cada equipo (Número de goles) se espera inferir el ganador de los encuentros.

Imágen de referencia a los diferentes modelos implementados



Random Forest

Continuando con la metodología propuesta para el SVM, se espera corroborar la precisión de dichos resultados con ayuda del Decision Tree al inferir el valor de una variable (Ganador de cada encuentro) en función de diversas características, como es el [Registro histórico de partidos ganados según el país](#) o el [Número de goles](#).



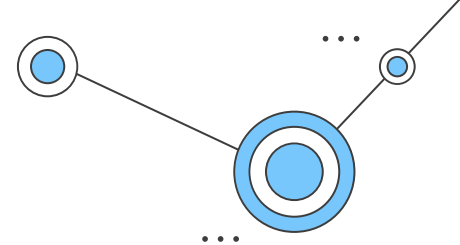


04

Resultados

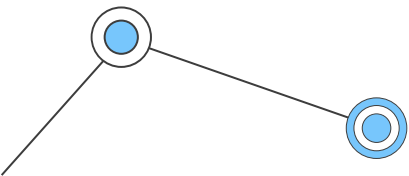


Posicionamiento de los equipos usando Regresión

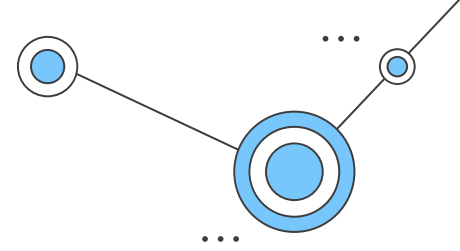


Grupo A	
Equipo	Puntos
Argentina	12
Chile	7
Paraguay	4
Bolivia	3
Uruguay	2

Grupo B	
Equipo	Puntos
Brasil	12
Colombia	9
Ecuador	4
Venezuela	2
Perú	1

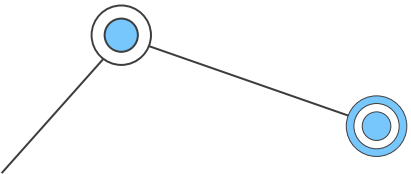


Posicionamiento de los equipos usando Clasificación

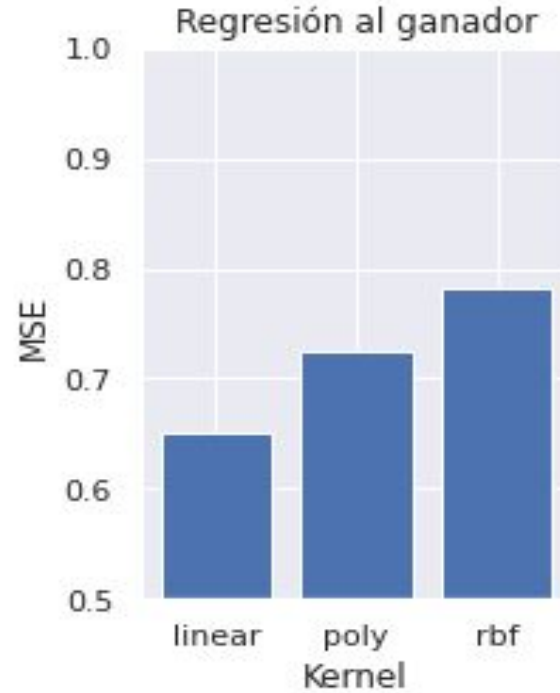
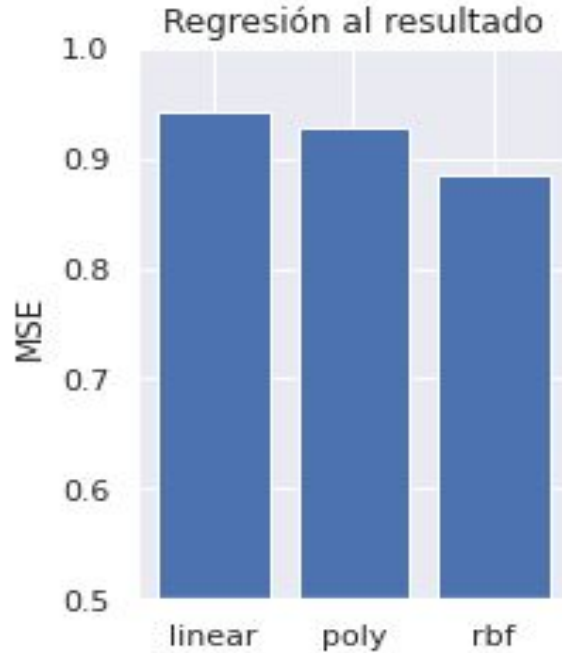


Grupo A	
Argentina	10
Uruguay	5
Chile	5
Paraguay	3
Bolivia	3

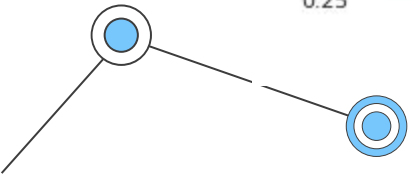
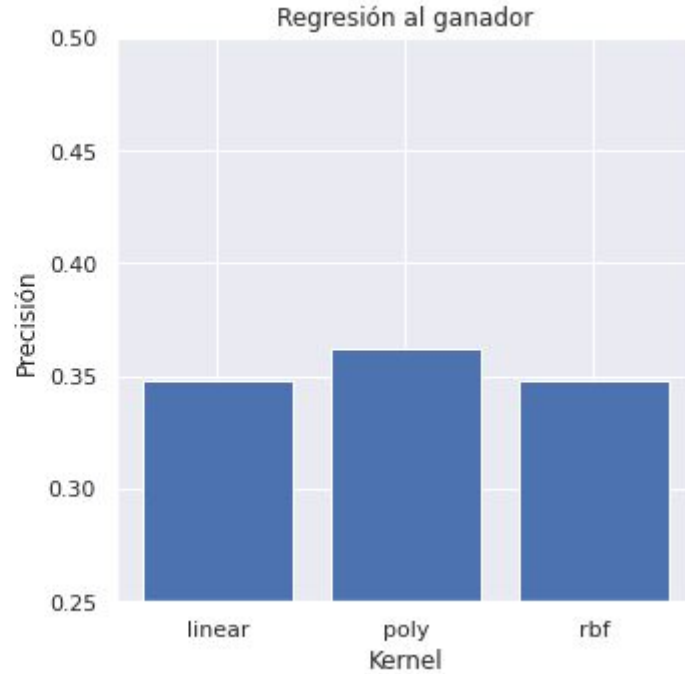
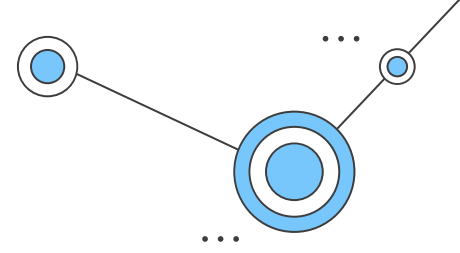
Grupo B	
Brasil	12
Colombia	6
Peru	5
Ecuador	2
Venezuela	2



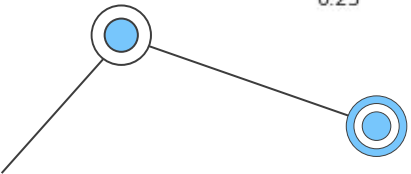
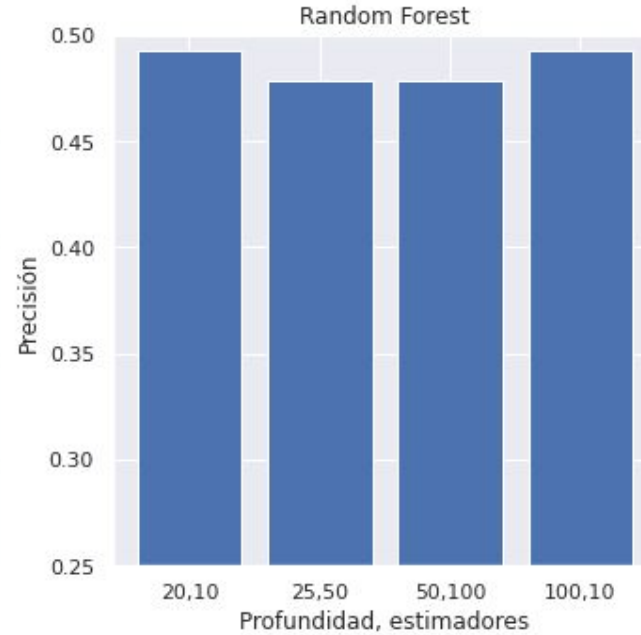
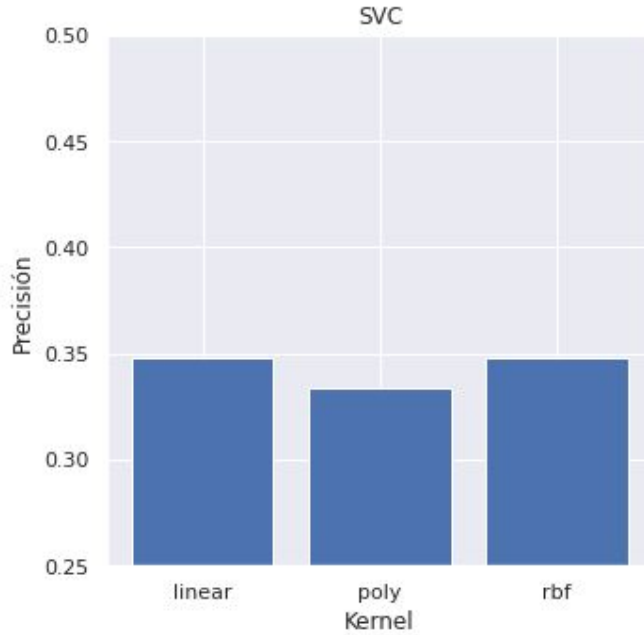
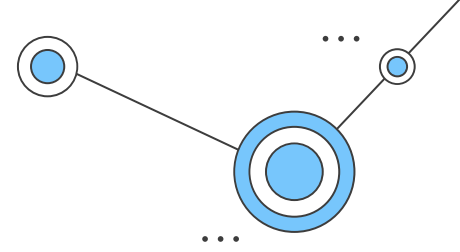
Desempeño de los métodos de regresión



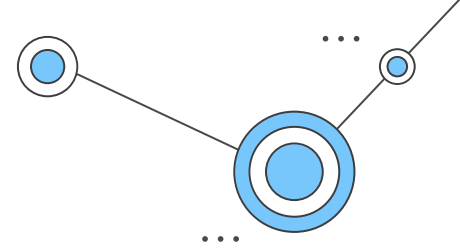
Precisión de los resultados de regresión



Desempeño de los métodos de clasificación



Resultados Copa América 2021



Brasil

Ganador Copa
América



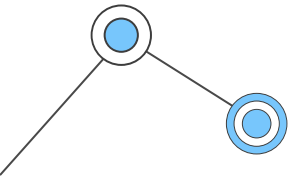
Argentina

Pierde partido contra
Brasil 2:0



Colombia

Gana partido contra
Ecuador por penales



Conclusiones

- El modelamiento a la hora de interpretar las inferencias realizadas mediante regresión jugó un papel importante para validar la premisa del trabajo expuesto
- El uso de distintos métodos de inferencia y clasificación como alternativa para disminuir la margen de error en cada modelo nos permitieron simplificar el alcance de nuestro proyecto



Gracias

