

Análisis y Predicción de Resultados en Partidas de Clash Royale

Trabajo Fin de Máster

Máster en Ciencia de Datos e Ingeniería de Datos en la Nube

Iván Fernández García

Junio, 2025



Tabla de contenidos (I)

01

Introducción

02

Creación
del Dataset

03

Análisis
Exploratorio
de Datos

04

Preprocesamiento
de Datos



Tabla de contenidos (II)

05

Modelado

06

Explicabilidad
y Despliegue

07

Conclusiones

08

Planificación
del Proyecto



01



Introducción



Motivación



- Aprovechar al máximo los datos disponibles
- Aplicar ML en un entorno competitivo real
- Complementar plataformas como *RoyaleAPI*



Crear una herramienta que nos permita tomar decisiones estratégicas y comprender cómo los diferentes factores del juego pueden influir en el desenlace de las partidas

Objetivos



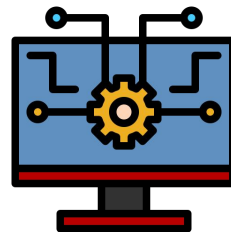
Crear un conjunto de datos a partir de la API de Clash Royale



Desarrollar un modelo capaz de predecir nuevas partidas



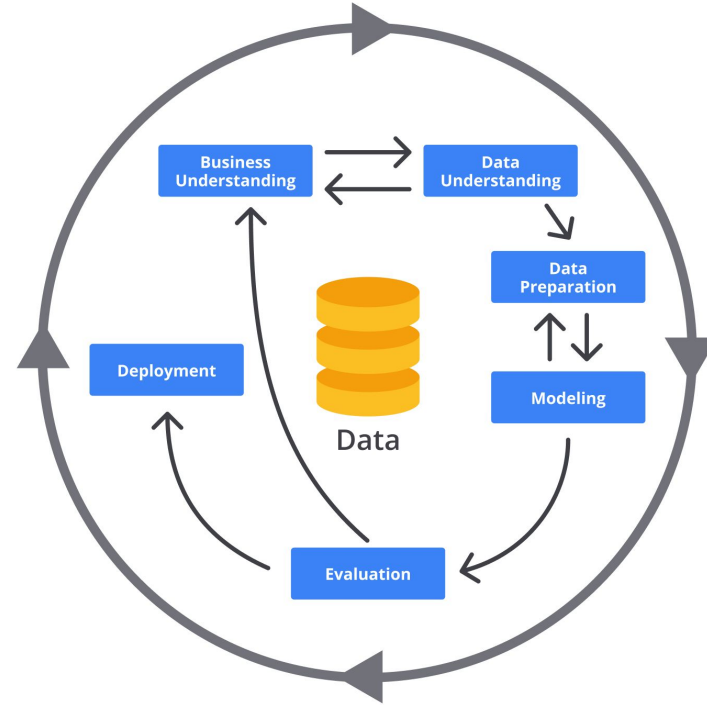
Evaluar los resultados y sacar conclusiones



Integrar el modelo en una aplicación web



Metodología



CRISP-DM (Cross-Industry Standard Process for Data Mining)



02

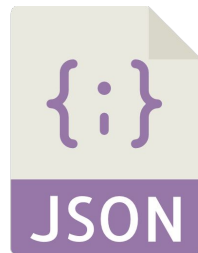


Creación del Dataset

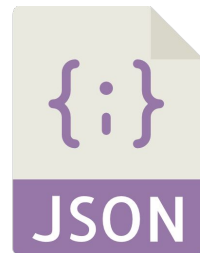
Adquisición de datos



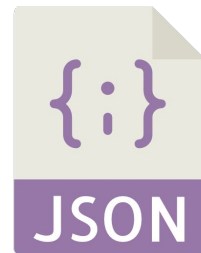
(API oficial)



Cartas



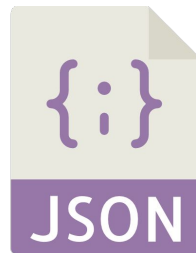
Cartas de
soporte



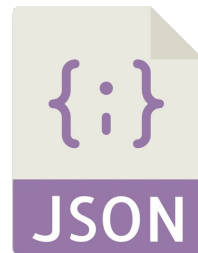
Partidas



(Otras fuentes)



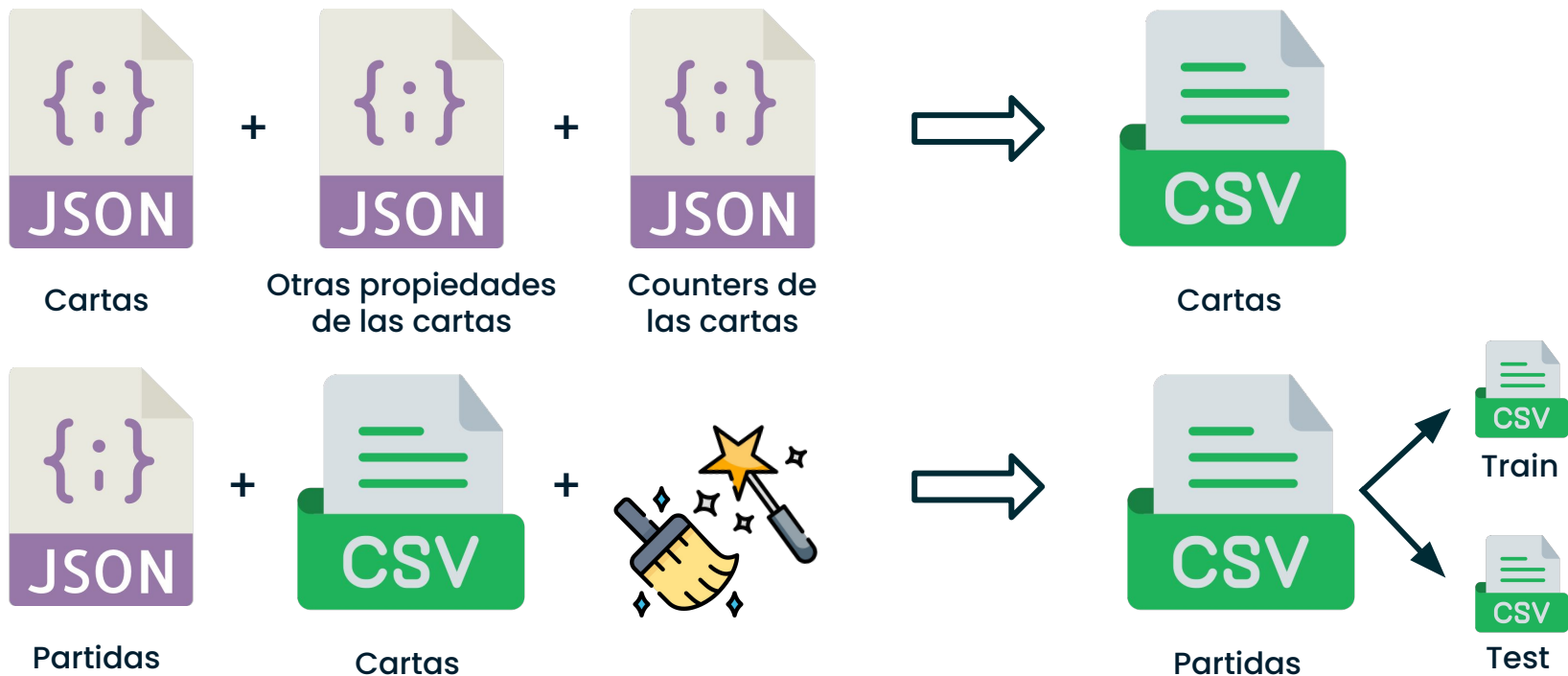
Otras propiedades
de las cartas



Counters de
las cartas



Limpieza y transformación





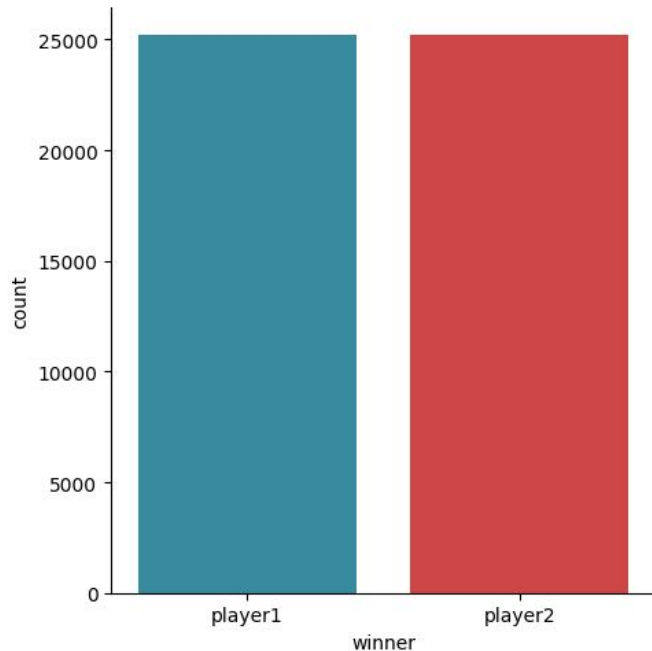
03



Análisis Exploratorio de Datos

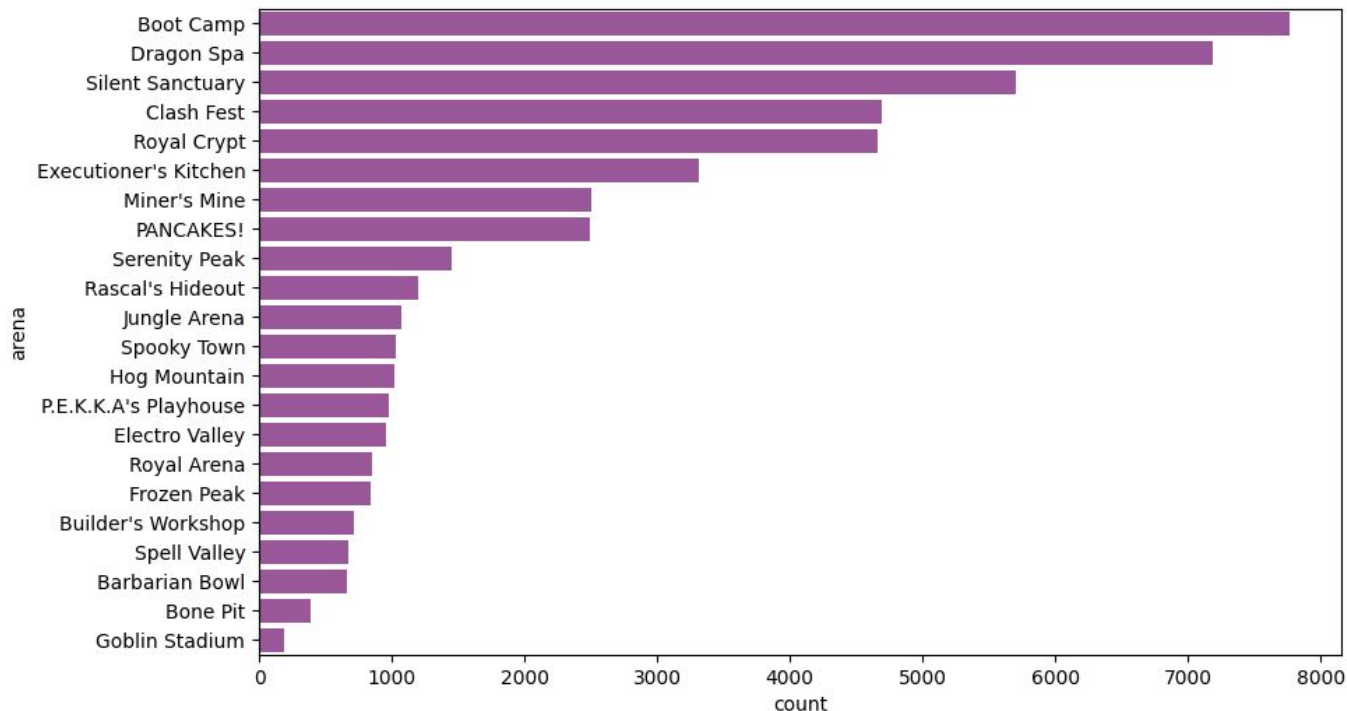


Distribución de la variable objetivo



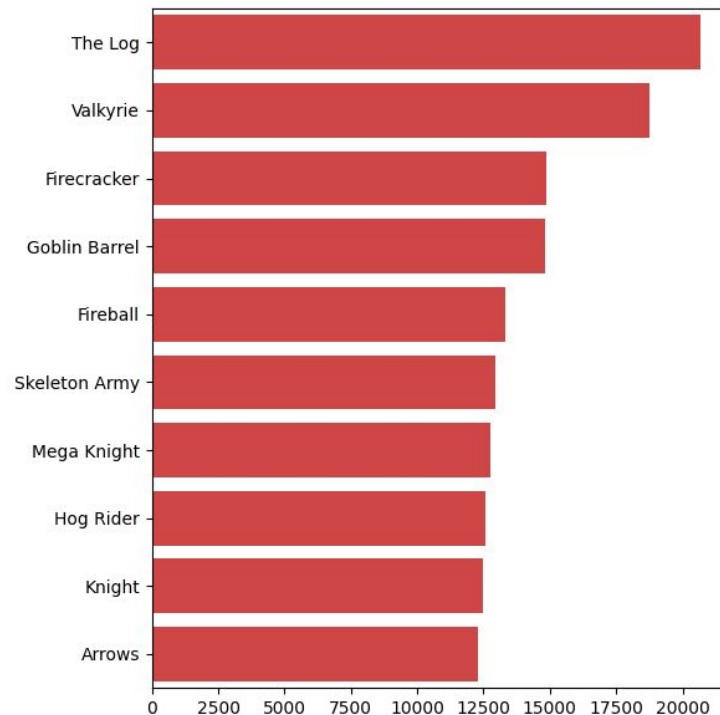
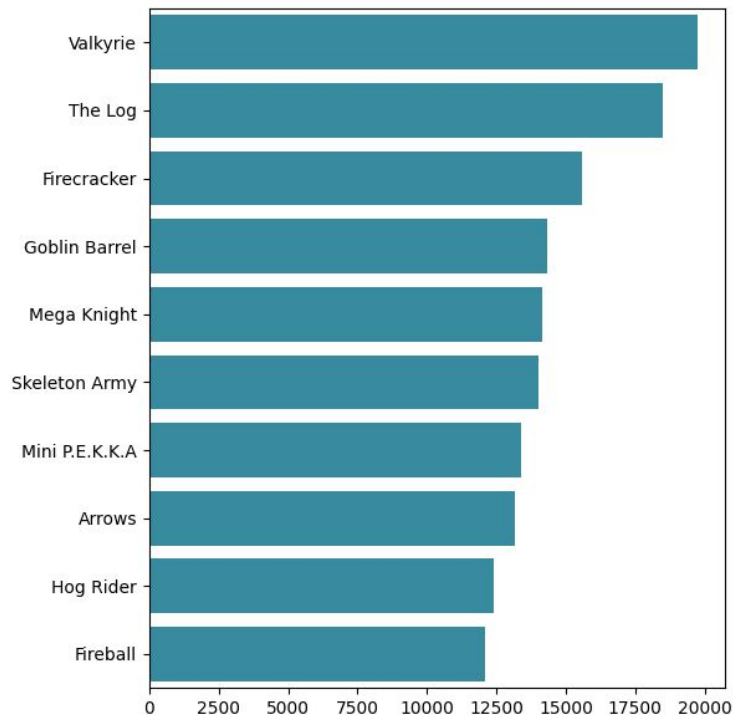


Partidas por arena



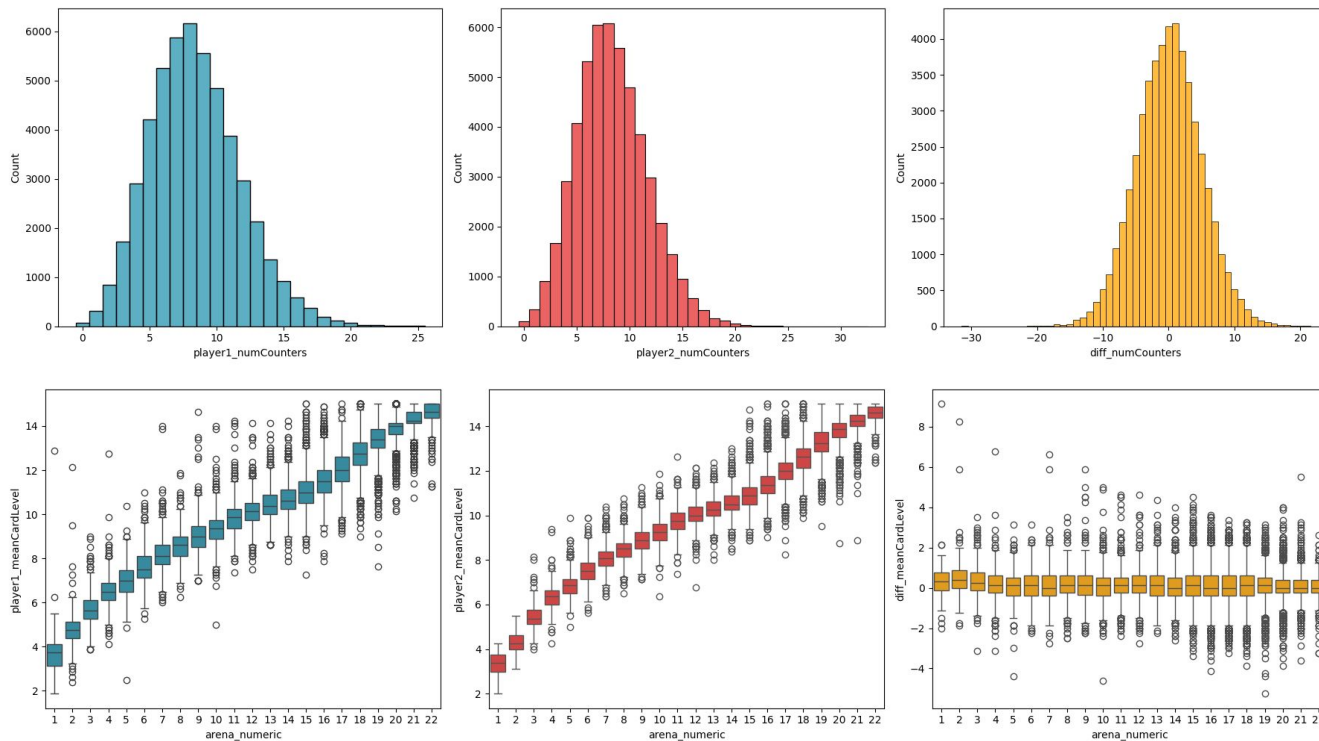


Cartas más utilizadas



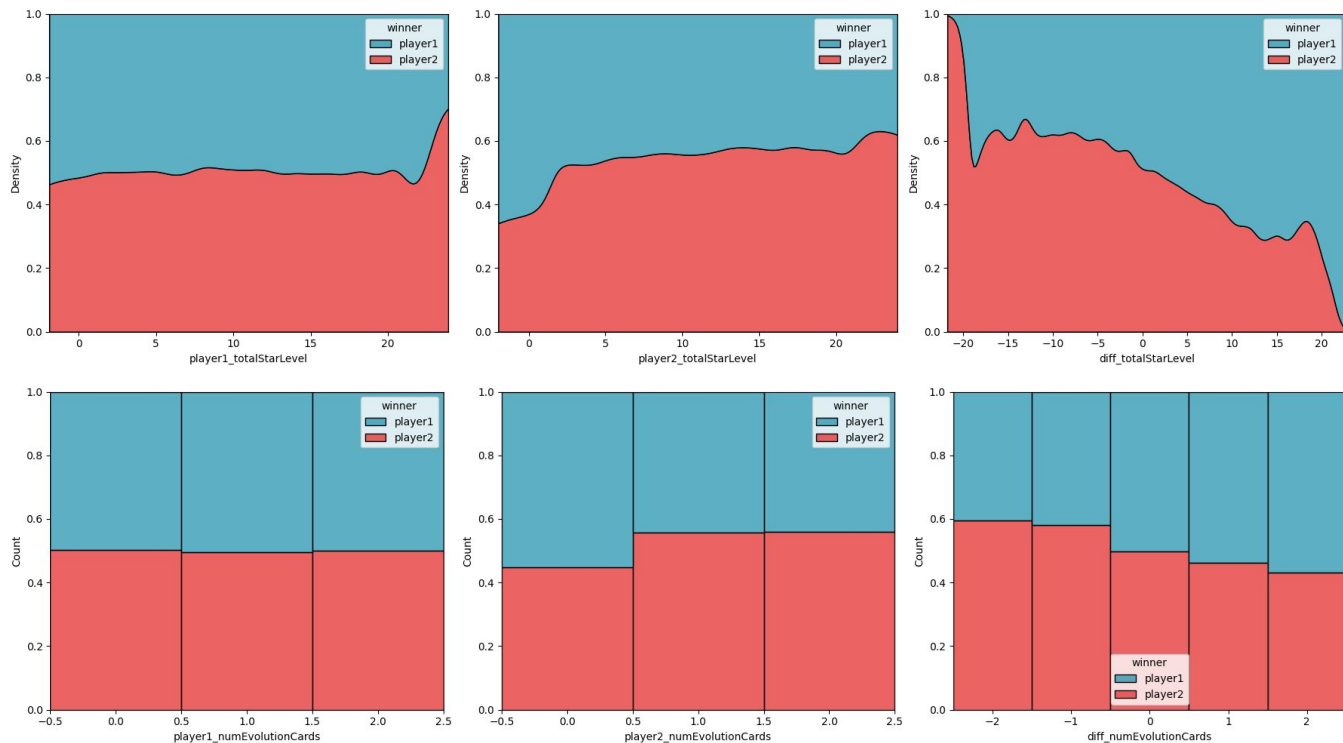


Análisis univariado y multivariado (I)



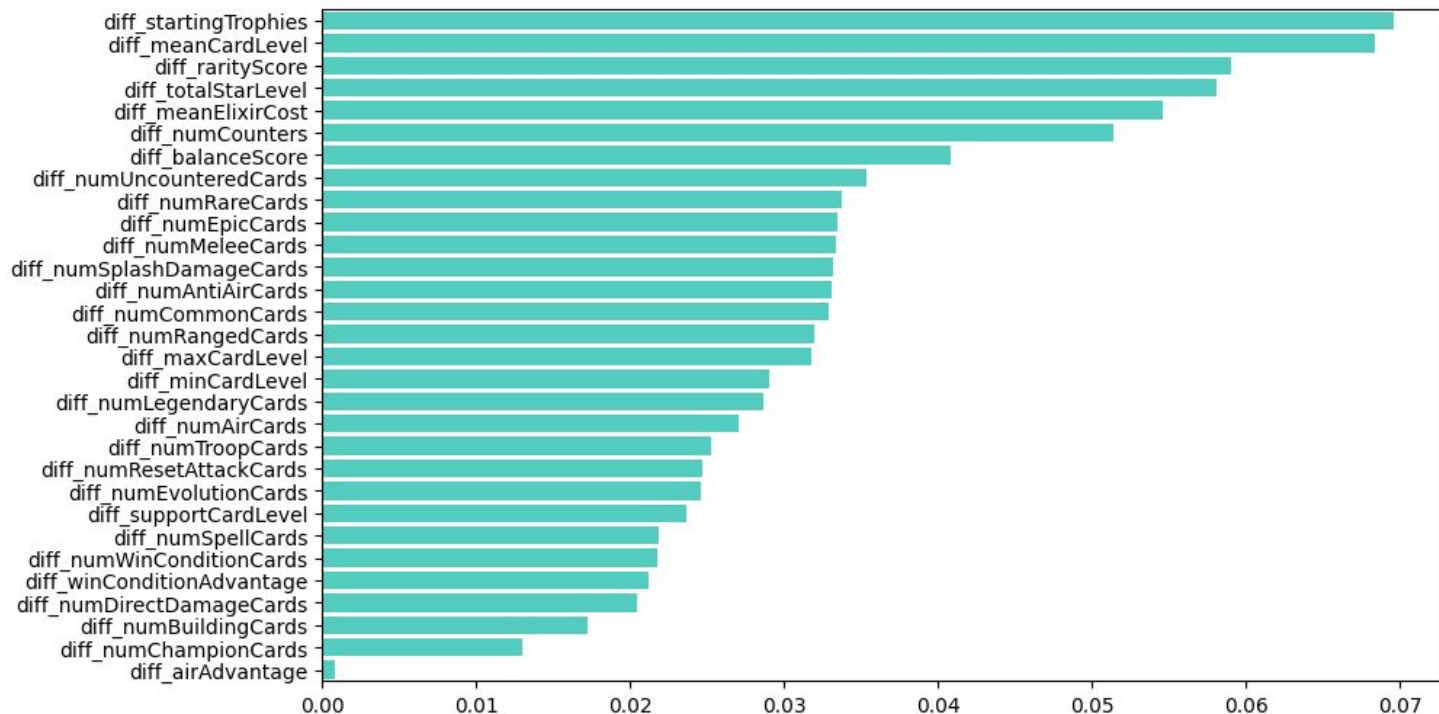


Análisis univariado y multivariado (II)





Importancia de las variables

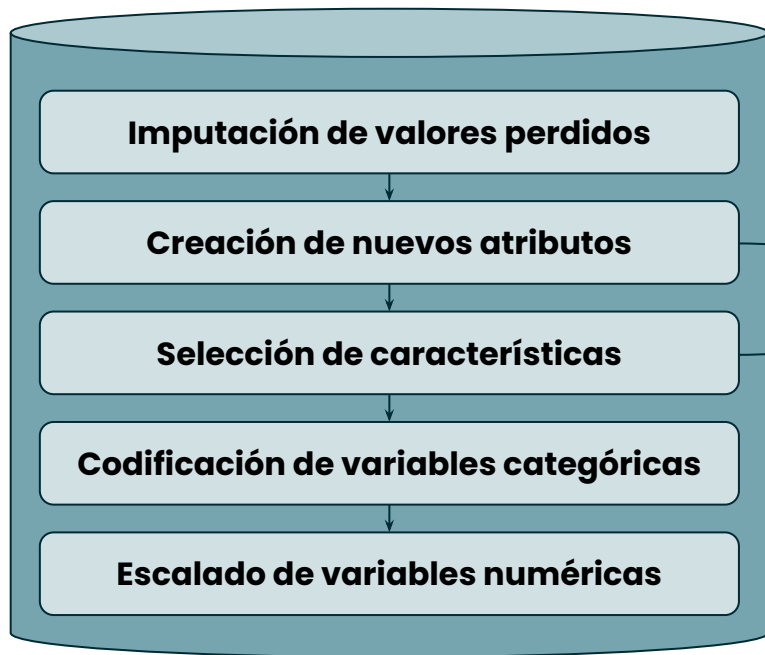




04 | Preprocesamiento de Datos



Pipelines de preprocesamiento



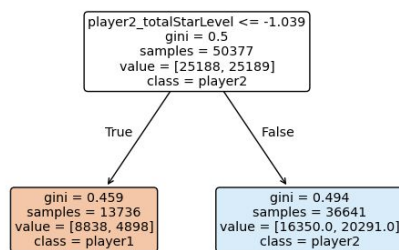
1. Binary + Numeric
2. Binary + Differences
3. Binary + Numeric + Differences
4. Only Differences
5. Selected Differences



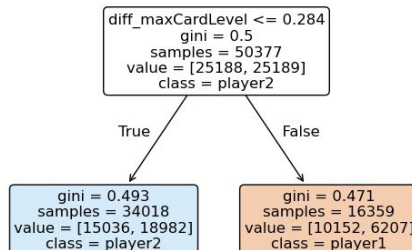
05 | Modelado



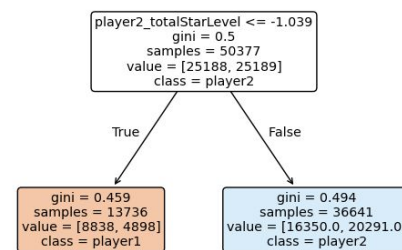
Baselines



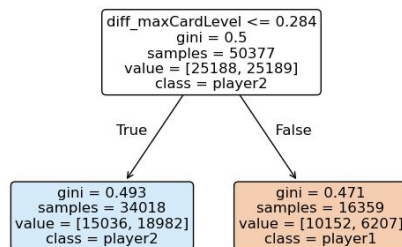
Binary + Numeric



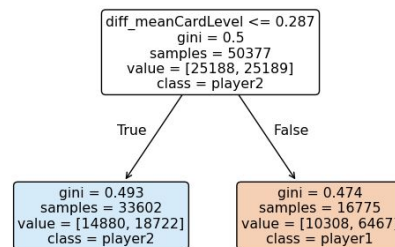
Binary + Differences



Binary + Numeric + Differences



Only Differences



Selected Differences



Validación y selección de modelos

- Combinar diferentes algoritmos de aprendizaje con cada pipeline de preprocesamiento
- Para cada combinación, optimizar los hiperparámetros mediante búsqueda en malla:
 - Definir varias configuraciones
 - Validar cada configuración mediante una 5-CV utilizando *accuracy* como métrica
 - Reentrenar el mejor modelo con todos los datos de entrenamiento

Modelos probados en total: **1330**

Modelos seccionados: **45** (mejor configuración por cada par *algoritmo-pipeline*)

Modelos para la evaluación final: **50** (los 45 modelos seleccionados y los 5 *baselines*)

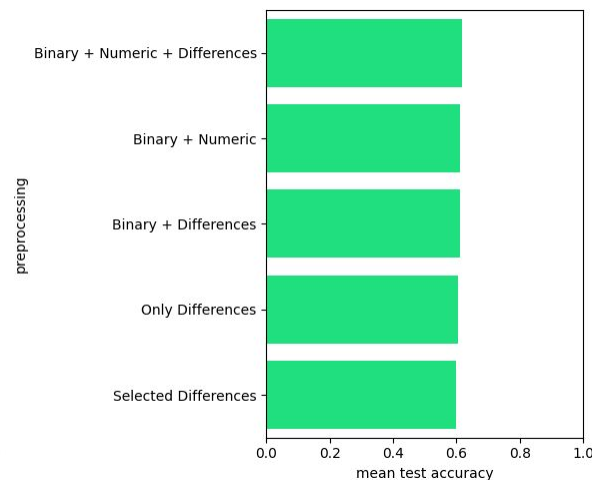
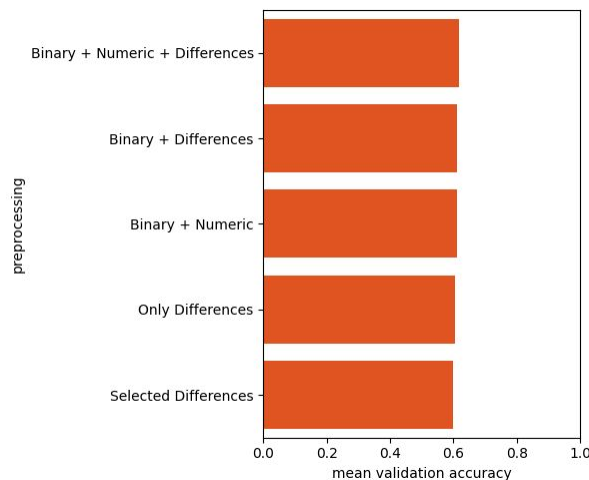
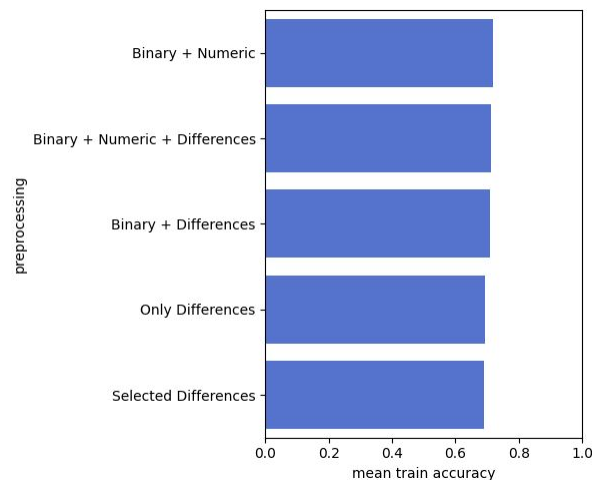


Evaluación final de modelos

model	train accuracy	val accuracy (10x5-CV)	test accuracy
Random Forest (Binary + Numeric + Differences)	0.994978	0.630036	0.630965
Gradient Boosting (Binary + Numeric + Differences)	0.654525	0.629686	0.628503
Histogram Gradient Boosting (Binary + Numeric + Differences)	0.672628	0.629631	0.628980
AdaBoost (Binary + Numeric + Differences)	0.711475	0.628916	0.625724
Gradient Boosting (Binary + Numeric)	0.719475	0.628160	0.630171
...
Baseline (Only Differences)	0.578319	0.578319	0.571417
Baseline (Binary + Numeric)	0.578220	0.578220	0.579198
Baseline (Binary + Numeric + Differences)	0.578220	0.577875	0.579198
Baseline (Selected Differences)	0.576255	0.576094	0.572052
KNN (Binary + Numeric)	1.000000	0.575937	0.580627

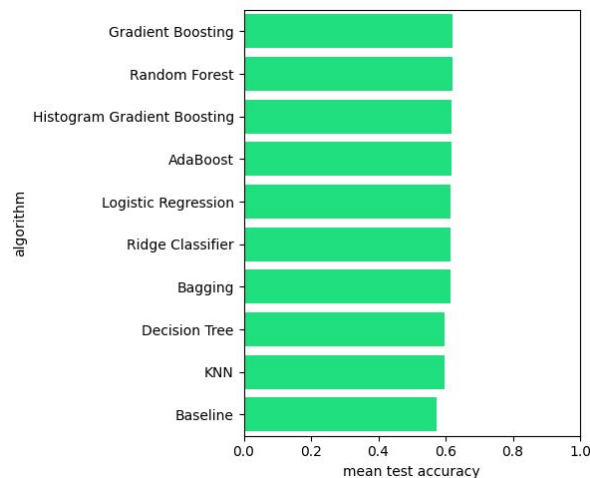
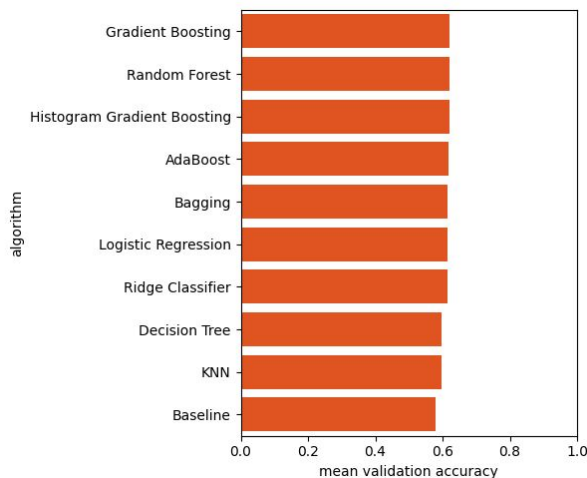
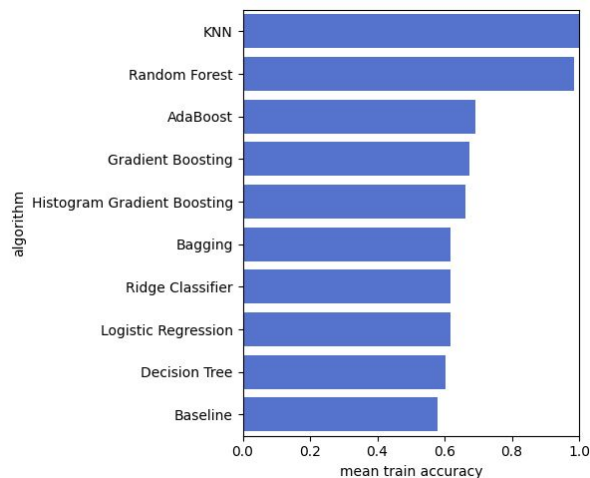


Rendimiento medio por preprocesamiento



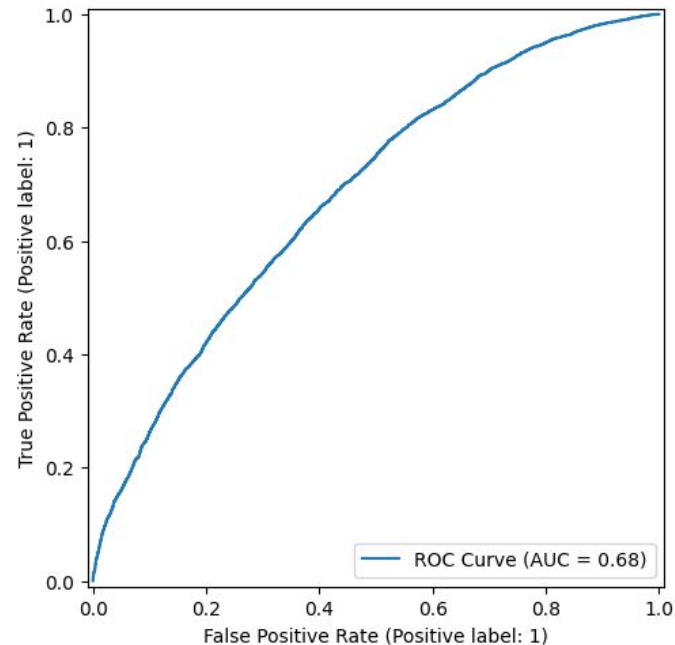
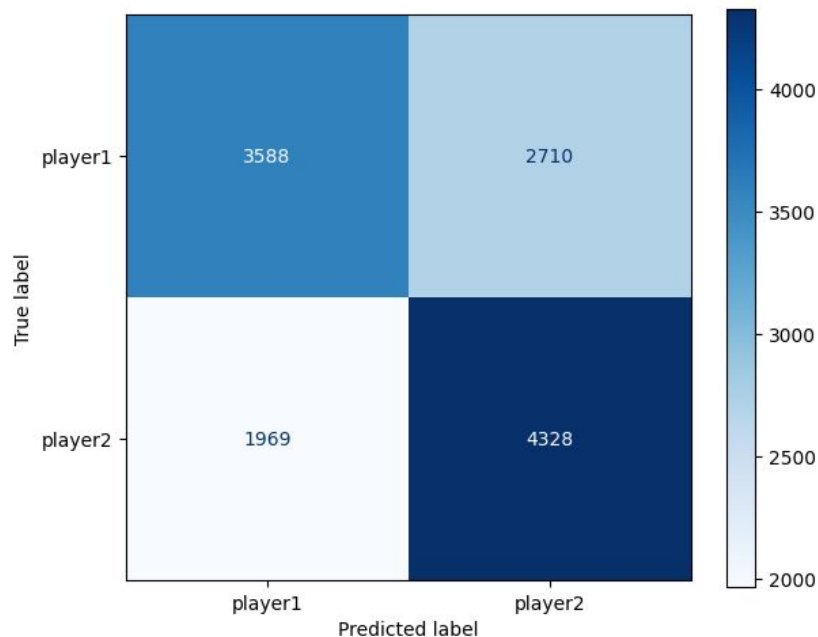


Rendimiento medio por algoritmo





Rendimiento del modelo sobre test





¿Cuánta incertidumbre hay?



count	mean	std	min	10%	50%	90%	max
12595	0.243361	0.180551	0.000034	0.039297	0.203618	0.517325	0.895580

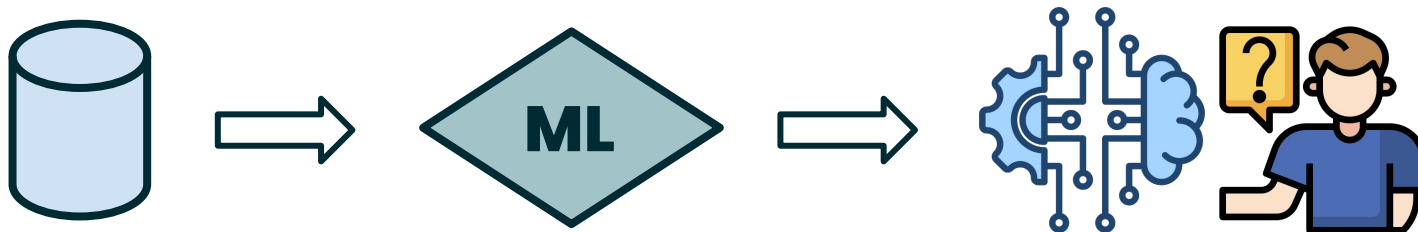
- Ninguna predicción alcanza una probabilidad del 95%
- 1 de cada 10 partidas se predicen con una probabilidad inferior al 52%
- 9 de cada 10 partidas se predicen con una probabilidad inferior al 76%
- Prácticamente la mitad de las partidas se predicen con una probabilidad inferior al 60%



06 | Explicabilidad y Despliegue



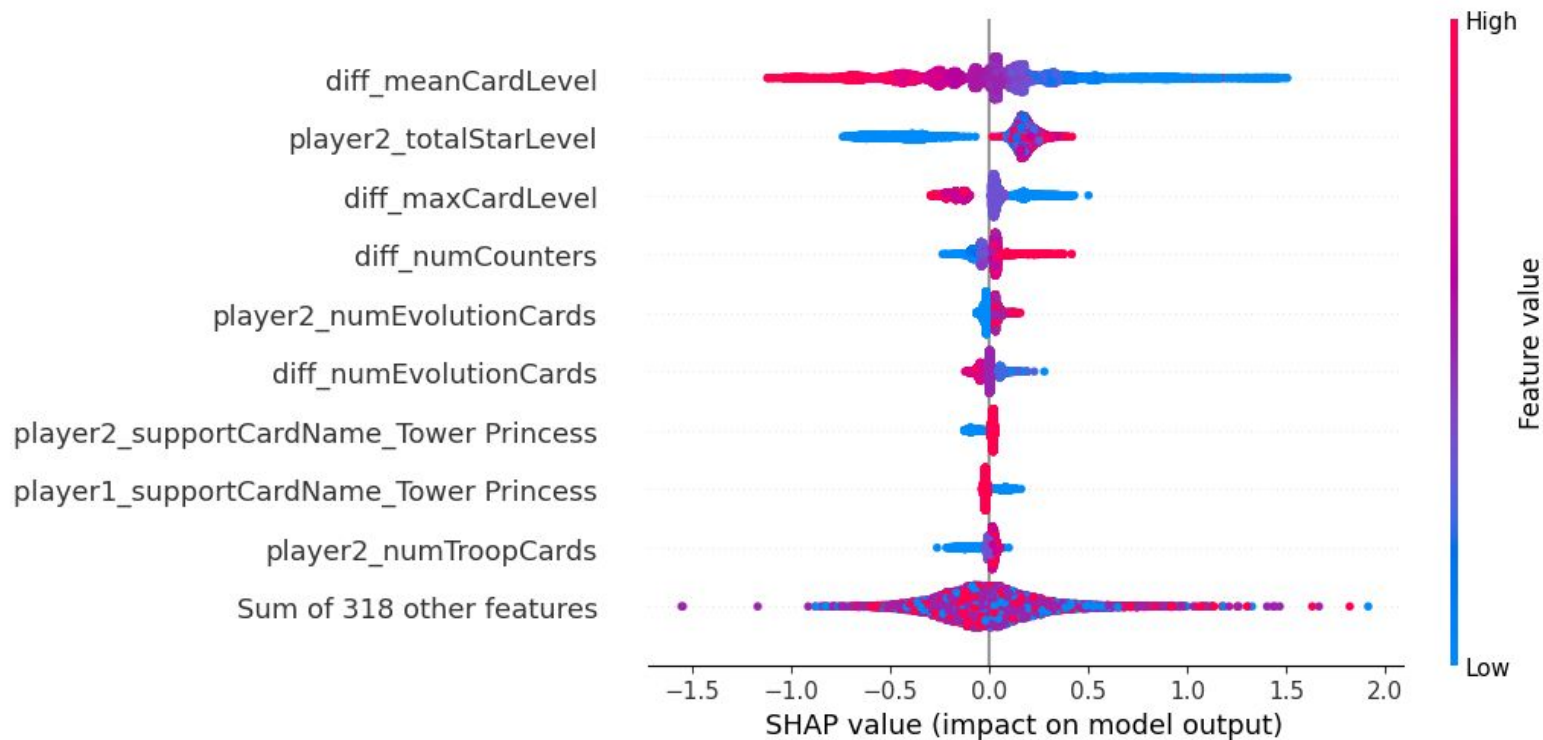
Explicabilidad



- ¿Qué variables son más determinantes en el resultado?
- ¿Cómo influyen los diferentes factores del juego en las predicciones?
- ¿Qué variables favorecen la victoria de cada jugador?
- ¿Cuáles pesan más y hacia qué lado se inclina la balanza?
- ¿Qué cambios serían necesarios para que el resultado fuera distinto?

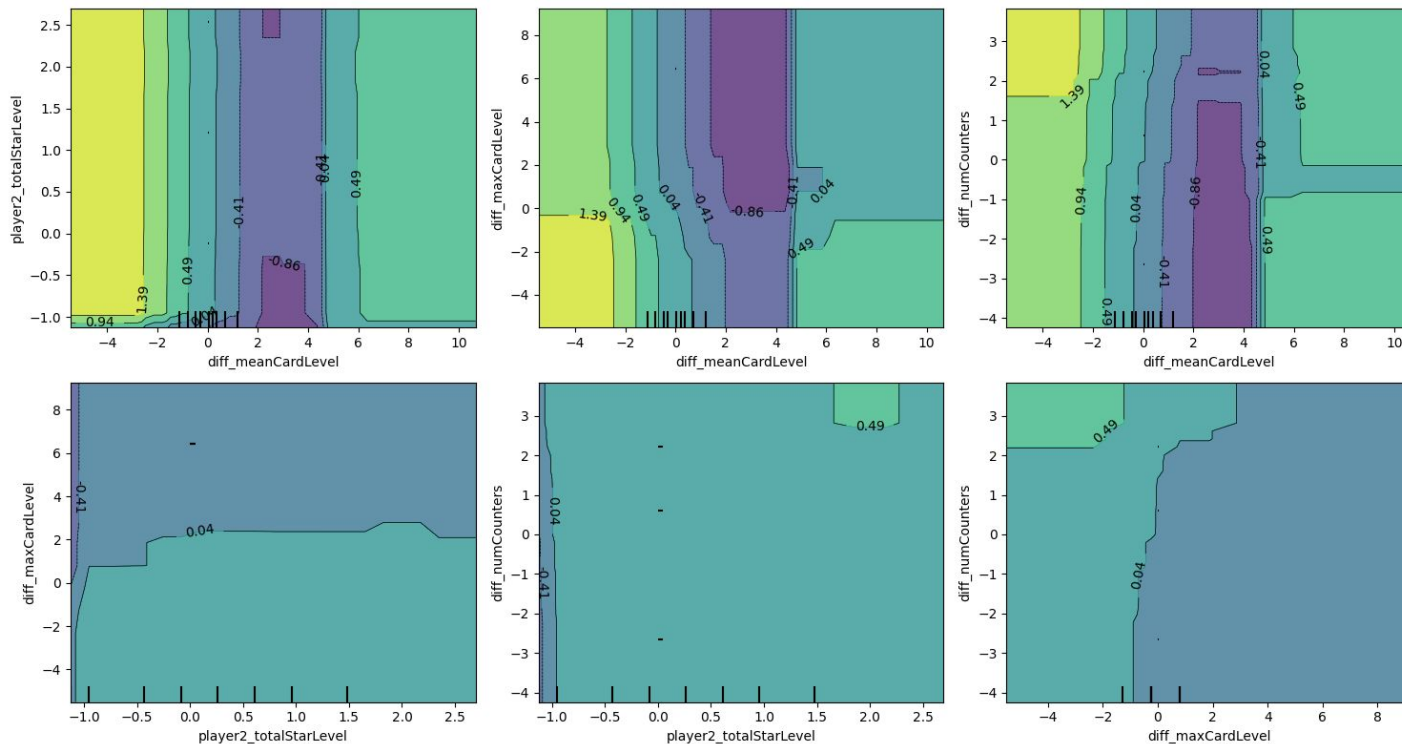


Interpretabilidad global: SHAP



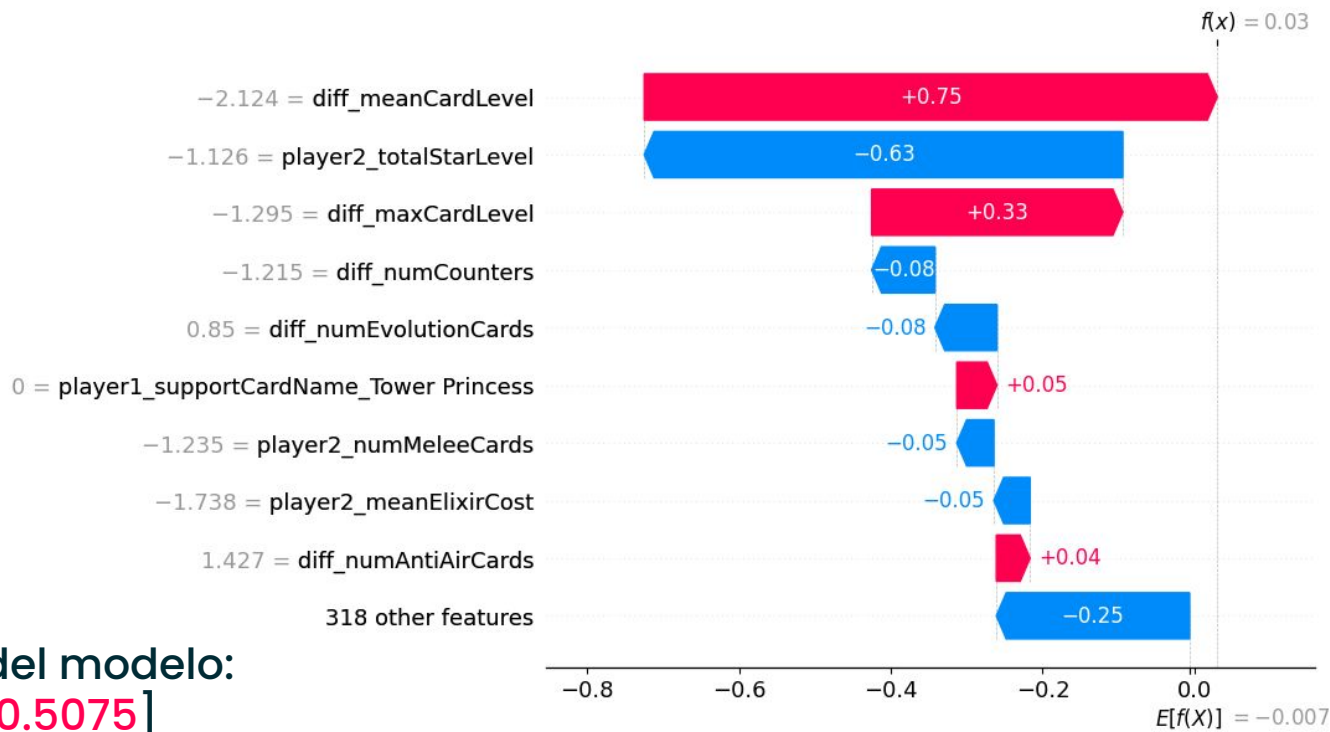


Interpretabilidad global: PDP





Interpretabilidad local: SHAP





Despliegue

Aplicación interactiva para predecir nuevas partidas:

- Cargar los perfiles de los jugadores a partir de sus tags
- Seleccionar los mazos de ambos jugadores
- Crear un nuevo registro con el formato adecuado
- Integrar el modelo y las técnicas de explicabilidad
- Mostrar el resultado de la predicción



Desarrollo de la aplicación (I)





Desarrollo de la aplicación (II)

Predicción de Resultados

#20ZGLQUU

Cargar jugadores


#88VYQQ6C

KILLER
★ Nivel 41
🏆 7682

XxByHabboxX
★ Nivel 36
🏆 5943

Predecir partida

Selecciona la tropa de las torres


Princesa de torre
Nivel 13


Cañonero
Nivel 6



Duquesa de dagas
Nivel 10

Selecciona la tropa de las torres



Princesa de torre
Nivel 12

Selecciona las cartas del mazo


Espíritu de fuego
(1)
Nivel 9


Recolector de elixir
(6)
Nivel 11


Baril de esqueletos
(3)
Nivel 11


Dragón infernal
(4)
Nivel 10

Selecciona las cartas del mazo


Esqueleto gigante
(6)
Nivel 8


Choza de bárbaros
(6)
Nivel 6


Duendes con lanza
(6)
Nivel 9


Espíritu de fuego
(1)
Nivel 9





Desarrollo de la aplicación (IV)





07 | Conclusiones



Conclusiones



Dominio del problema



Complejidad e
incertidumbre



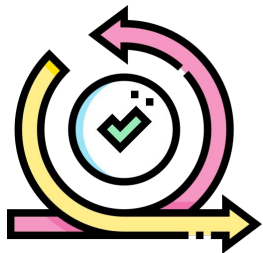
Importancia
de los niveles



Matchmaking



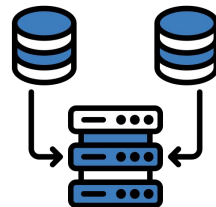
Futuras mejoras



Nuevas iteraciones



Más datos



Información
de los perfiles



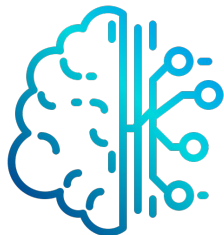
Automatización del
flujo y mantenimiento



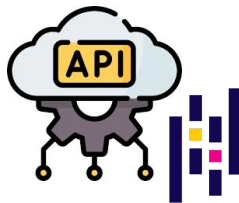
Modelos para otros
modos de juego



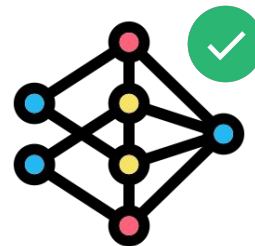
Competencias desarrolladas



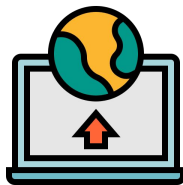
Competencias de
computación



Adquisición y
manipulación
de datos



Preprocesamiento
y modelado



Implementación de
una solución práctica



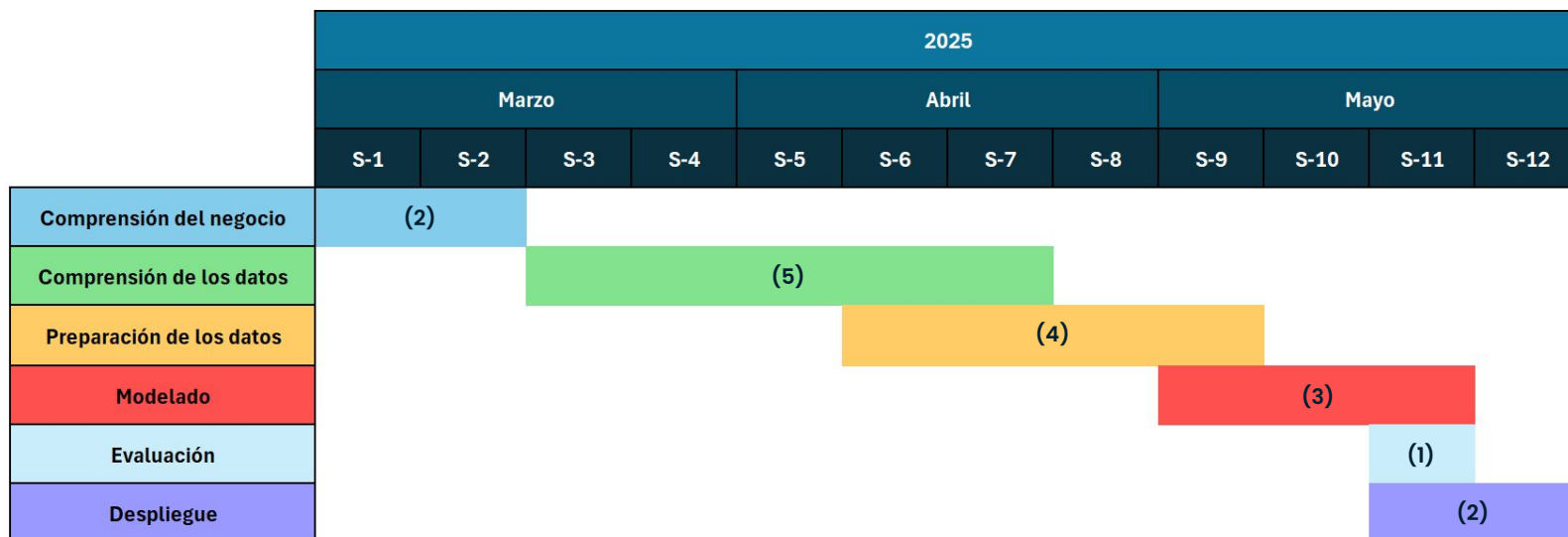
Técnicas de
explicabilidad



08 | Planificación del Proyecto



Planificación del proyecto





¡Gracias por su atención!

¿Alguna pregunta?