



ANÁLISIS PLANTILLA RMA LA LIGA 21-25



OBJETIVOS

- Análisis de rendimiento por conjunto de posiciones para evaluación de cambio de jugadores.
- Rendimiento y correlación de edad-tiempo jugado comprendiendo qué jugadores podrían mantenerse a futuro , quienes se podrían jubilar pronto y la pérdida de ese puesto
- Evolución de la plantilla por torneo, análisis en rotación de jugadores
- Comparativa reflexiva con la liga de la temporada siguiente a la salida de Carlos Ancelotti (25-26). Predicción de tarjetas amarillas, asistencias, goles-asistencias, goles.

DATOS



Se utilizan 4 tablas cada una de una temporada distinta de la Liga, donde se resume las estadísticas estándares por jugador



La fuente de los datos es FBref.
Organización dedicada a la recolección de estadísticas en el mundo del fútbol.

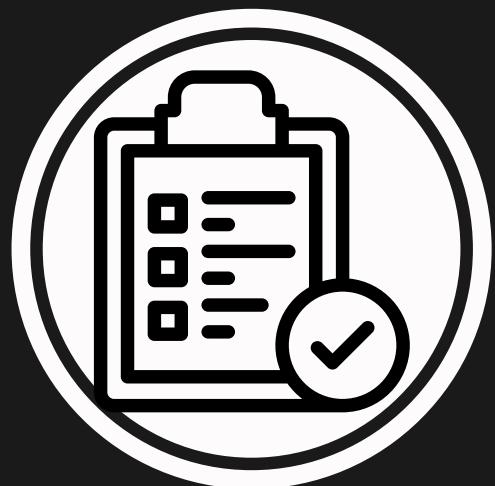
			Playing Time				Performance							Expected			Progression			Per 90 Minutes													
Player	Nation	Pos	Age	MP	Starts	Min	90s	Gls	Ast	G+A	G-PK	PK	PkAtt	CrdY	CrdR	xG	npxG	xAG	npxG+xAG	PrgC	PrgP	PrgR	Gls	Ast	G+A	G-PK	G+A-PK	xG	xAG	xG+xAG	npxG	npxG+xAG	Matches
Federico Valverde	URU	MF,DF	26	36	34	3,032	33.7	6	4	10	6	0	0	4	0	2.7	2.7	3.1	5.8	66	226	76	0.18	0.12	0.30	0.18	0.30	0.08	0.09	0.17	0.08	0.17	Matches
Kylian Mbappé	FRA	FW	25	34	34	2,907	32.3	31	3	34	24	7	9	3	1	25.9	18.6	7.7	26.3	150	140	386	0.96	0.09	1.05	0.74	0.84	0.80	0.24	1.04	0.58	0.81	Matches
Aurélien Tchouaméni	FRA	DF,MF	24	32	31	2,689	29.9	0	0	0	0	0	0	5	0	1.7	1.7	0.6	2.2	19	154	13	0.00	0.00	0.00	0.00	0.00	0.06	0.02	0.07	0.06	0.07	Matches
Thibaut Courtois	BEL	GK	32	30	30	2,700	30.0	0	0	0	0	0	0	1	0	0.0	0.0	0.1	0.1	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Matches
Jude Bellingham	ENG	MF	21	31	29	2,488	27.6	9	8	17	8	1	2	5	1	11.4	9.9	3.9	13.7	74	199	143	0.33	0.29	0.61	0.29	0.58	0.41	0.14	0.55	0.36	0.50	Matches
Antonio Rüdiger	GER	DF	31	29	26	2,291	25.5	0	0	0	0	0	0	3	0	1.7	1.7	0.5	2.3	20	94	4	0.00	0.00	0.00	0.00	0.00	0.07	0.02	0.09	0.07	0.09	Matches
Lucas Vázquez	ESP	DF	33	32	25	2,189	24.3	1	5	6	1	0	0	4	0	1.4	1.4	4.7	6.1	37	99	194	0.04	0.21	0.25	0.04	0.25	0.06	0.19	0.25	0.06	0.25	Matches
Vinicius Júnior	BRA	FW	24	30	24	2,253	25.0	11	8	19	9	2	3	8	1	10.4	8.2	8.0	16.2	169	85	256	0.44	0.32	0.76	0.36	0.68	0.41	0.32	0.73	0.33	0.65	Matches
Fran García	ESP	DF	24	31	24	2,171	24.1	0	2	2	0	0	0	0	0	0.9	0.9	1.8	2.7	78	119	177	0.00	0.08	0.08	0.00	0.08	0.04	0.08	0.11	0.04	0.11	Matches
Rodrygo	BRA	FW,MF	23	30	22	1,928	21.4	6	5	11	6	0	0	0	0	4.0	3.9	3.0	6.9	113	108	253	0.28	0.23	0.51	0.28	0.51	0.19	0.14	0.33	0.18	0.32	Matches
Raúl Asencio	ESP	DF	21	23	18	1,670	18.6	0	1	1	0	0	0	3	0	0.1	0.1	0.6	0.7	5	27	4	0.00	0.05	0.05	0.00	0.05	0.01	0.03	0.04	0.01	0.04	Matches
Luka Modrić	CRO	MF	38	35	17	1,827	20.3	2	6	8	2	0	0	7	0	1.6	1.6	7.0	8.5	62	208	63	0.10	0.30	0.39	0.10	0.39	0.08	0.34	0.42	0.08	0.42	Matches
Brahim Díaz	MAR	FW,MF	24	31	16	1,385	15.4	4	2	6	4	0	0	0	0	5.3	5.3	3.3	8.6	39	74	110	0.26	0.13	0.39	0.26	0.39	0.34	0.22	0.56	0.34	0.56	Matches
Dani Ceballos	ESP	MF	27	23	15	1,215	13.5	0	0	0	0	0	0	4	0	0.4	0.4	1.6	2.0	42	150	17	0.00	0.00	0.00	0.00	0.00	0.03	0.12	0.15	0.03	0.15	Matches
Arda Güler	TUR	MF,FW	19	28	14	1,250	13.9	3	4	7	3	0	0	1	0	3.3	3.3	4.5	7.7	34	93	86	0.22	0.29	0.50	0.22	0.50	0.24	0.32	0.56	0.24	0.56	Matches
Ferland Mendy	FRA	DF	29	14	12	1,005	11.2	0	1	1	0	0	0	1	1	0.0	0.0	0.5	0.5	12	36	40	0.00	0.09	0.09	0.00	0.09	0.00	0.05	0.05	0.00	0.05	Matches
Eduardo Camavinga	FRA	MF,DF	21	19	11	1,102	12.2	1	2	3	1	0	0	2	0	0.4	0.4	0.8	1.2	26	61	31	0.08	0.16	0.25	0.08	0.25	0.04	0.06	0.10	0.04	0.10	Matches
Éder Militão	BRA	DF	26	12	11	938	10.4	1	1	2	1	0	0	2	0	0.6	0.6	0.5	1.1	7	46	1	0.10	0.10	0.19	0.10	0.19	0.06	0.05	0.11	0.06	0.11	Matches
Andriy Lunin	UKR	GK	25	7	7	630	7.0	0	1	1	0	0	0	0	0	0.0	0.0	0.9	0.9	0	0	0	0.00	0.14	0.14	0.00	0.14	0.00	0.13	0.00	0.13	0.00	Matches
Dani Carvajal	ESP	DF	32	8	7	609	6.8	1	0	1	1	0	0	1	0	1.3	1.3	0.2	1.5	13	27	48	0.15	0.00	0.15	0.15	0.15	0.19	0.03	0.22	0.19	0.22	Matches
David Alaba	AUT	DF	32	7	4	317	3.5	0	0	0	0	0	0	1	0	0.0	0.0	0.1	0.1	2	7	4	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.02	Matches
Endrick	BRA	FW	18	22	3	363	4.0	1	0	1	1	0	0	1	0	2.5	2.5	0.5	3.0	12	8	32	0.25	0.00	0.25	0.25	0.25	0.61	0.12	0.73	0.61	0.73	Matches
Jacobo Ramón	ESP	DF	19	3	2	204	2.3	1	0	1	1	0	0	1	0	0.2	0.2	0.2	0.4	0	1	2	0.44	0.00	0.44	0.44	0.44	0.07	0.09	0.16	0.07	0.16	Matches
Jesús Vallejo	ESP	DF	27	4	1	138	1.5	0	1	1	0	0	0	0	0	0.0	0.0	0.1	0.1	0	5	1	0.00	0.65	0.65	0.00	0.65	0.00	0.08	0.00	0.08	0.08	Matches
Fran	ESP	GK	19	1	1	90	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0.00	0.00									

PREDICCIONES



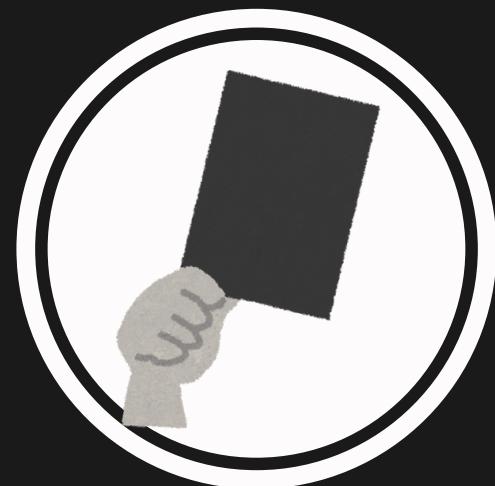
GOLES

Regresión



ASISTENCIAS

Regresión



TARJETAS

Clasificación



RENDIMIENTO

Regresión

POTENCIALES MODELOS DE ML

Se proponen inicialmente estos modelos, pero dependiendo de la distribución de datos, ajuste y rendimiento se pueden modificar para resultados más apropiados

1

REGRESIÓN LINEAL

Predicción de asistencias, goles, tarjetas

2

RANDOM FOREST

Predicción de rendimiento/minutos en partido por jugador

3

XGBOOST

Nos sirve para el efecto combinado de los pases progresivos y el xG

ESTOS ANÁLISIS
DEBERÁN SER
INTERPRETADOS
CONSIDERANDO LAS
DISTINTAS POSICIONES
DE LOS JUGADORES

MÉTRICAS DE EVALUACIÓN

Regresion lineal

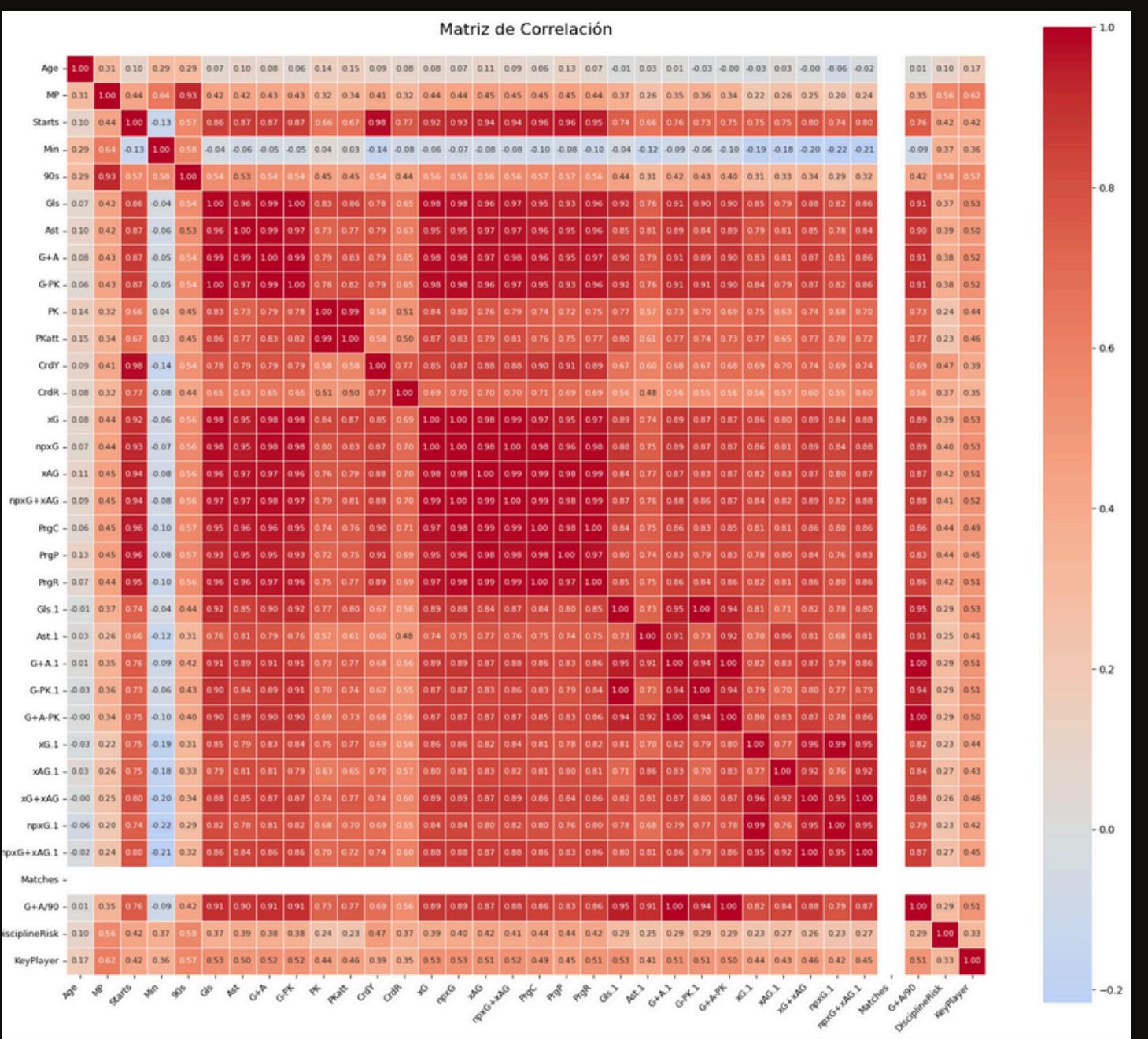
- Error Absoluto medio (MAE)
- Error cuadratico medio (MSE)
- Coeficiente de determinacion (R^2)
- Error Porcentual Absoluto Medio (MAPE)
- etc

Random Forest, XGboost y modelos de clasificación:

- Accuracy
- Precision
- Recall
- F1-Score
- etc

¿CÓMO RESULTARON LOS MODELOS?

Comentarios : Por la alta correlación de las variables normales y las divididas por 90. Es que solo se trabaja con una de las mitades. Luego se separan los datos dejando en los predictores aquellas entradas de las temporadas 21-24. Se dejan en las features que no tienen relaciones directas.



MODELOS DE REGRESIÓN

Regresión Lineal Múltiple

Predicción de Goles

```
=====  
REGRESION LINEAL MULTIPLE - Prediccion de Goles (G)  
=====
```

```
Features utilizadas (12): ['Age', 'MP', 'Starts', 'Min', '90s', '  
'
```

Resultados en Entrenamiento (2021-2024):

R2 Score: 0.9660

RMSE: 2.8327

MAE: 2.0052

Resultados en Test (2024-2025):

R2 Score: 0.9609

RMSE: 3.0819

MAE: 1.9705

Predicción de Asistencias

```
=====  
REGRESION LINEAL MULTIPLE - Prediccion de Asistencias (A)  
=====
```

```
Features utilizadas (12): ['Age', 'MP', 'Starts', 'Min', '90s', 'PK', 'PKatt', 'CrdY'  
'
```

Resultados en Entrenamiento (2021-2024):

R2 Score: 0.9785

RMSE: 1.6405

MAE: 1.1879

Resultados en Test (2024-2025):

R2 Score: 0.9404

RMSE: 2.5643

MAE: 1.6732

O Regresión Lineal Logística

REGRESION LOGISTICA - Clasificacion de DisciplineRisk

Features utilizadas (10): ['Age', 'MP', 'Starts', 'Min', '90s', 'PK', 'PKatt', 'PrgC'

Resultados en Entrenamiento (2021-2024):

Accuracy: 0.7753

Resultados en Test (2024-2025):

Accuracy: 0.7742

Reporte de Clasificacion (Test):

	precision	recall	f1-score	support
No Risk	0.86	0.82	0.84	22
Risk	0.60	0.67	0.63	9
accuracy			0.77	31
macro avg	0.73	0.74	0.73	31
weighted avg	0.78	0.77	0.78	31

Matriz de Confusion:

```
[[18  4]
 [ 3  6]]
```

REGRESION LOGISTICA - Clasificacion de KeyPlayer

Features utilizadas (12): ['Age', 'MP', 'Starts', 'Min', '90s', 'PK', 'PKatt', 'CrdY'

Resultados en Entrenamiento (2021-2024):

Accuracy: 0.9663

Resultados en Test (2024-2025):

Accuracy: 0.9032

Reporte de Clasificacion (Test):

	precision	recall	f1-score	support
Not Key	0.90	0.95	0.93	20
Key Player	0.90	0.82	0.86	11
accuracy			0.90	31
macro avg	0.90	0.88	0.89	31
weighted avg	0.90	0.90	0.90	31

Matriz de Confusion:

```
[[19  1]
 [ 2  9]]
```

Clasificación conducta

Clasificación de Key Player

Xgboost classifier

XGBOOST CLASSIFIER - Clasificacion de DisciplineRisk

Features utilizadas (10): ['Age', 'MP', 'Starts', 'Min', '90s', 'PK', 'PKatt', 'PrgC'

Resultados en Entrenamiento (2021-2024):

Accuracy: 1.0000

Resultados en Test (2024-2025):

Accuracy: 0.7742

Reporte de Clasificacion (Test):

	precision	recall	f1-score	support
No Risk	0.83	0.86	0.84	22
Risk	0.62	0.56	0.59	9
accuracy			0.77	31
macro avg	0.73	0.71	0.72	31
weighted avg	0.77	0.77	0.77	31

Matriz de Confusion:

```
[[19  3]
 [ 4  5]]
```

Clasificación conducta

XGBOOST CLASSIFIER - Clasificacion de KeyPlayer

Features utilizadas (12): ['Age', 'MP', 'Starts', 'Min', '90s', 'PK', 'PKatt', 'CrdY'

Resultados en Entrenamiento (2021-2024):

Accuracy: 1.0000

Resultados en Test (2024-2025):

Accuracy: 0.9355

Reporte de Clasificacion (Test):

	precision	recall	f1-score	support
Not Key	0.95	0.95	0.95	20
Key Player	0.91	0.91	0.91	11
accuracy			0.94	31
macro avg	0.93	0.93	0.93	31
weighted avg	0.94	0.94	0.94	31

Matriz de Confusion:

```
[[19  1]
 [ 1 10]]
```

Clasificación de Key Player

¡CONCLUSIONES!

Tuvimos un buen rendimiento en los modelos, tanto XGBoost como la regresión logística tuvieron buenos resultados en precisión.

El único inconveniente fue que para el modelo de regresión lineal múltiple nos dio un RMSE alto, el cual nos indica que falla en 3 goles, lo cual bajo nuestro punto de vista no es aceptable.

En XGBoost muestra una tendencia a overfitting en el set de entrenamiento pero su rendimiento de prueba es muy preciso.

Por otra parte DisciplineRisk tuvo un Accuracy de 0.7742, lo cual nos señala que la conducta disciplinaria depende de factores no capturados en las metricas de rendimiento y juego utilizadas.

MUCHAS GRACIAS

Integrantes:

Vicente Muñoz

Francisca Yepsen

Pilar Vallotton

Cristóbal Soto

