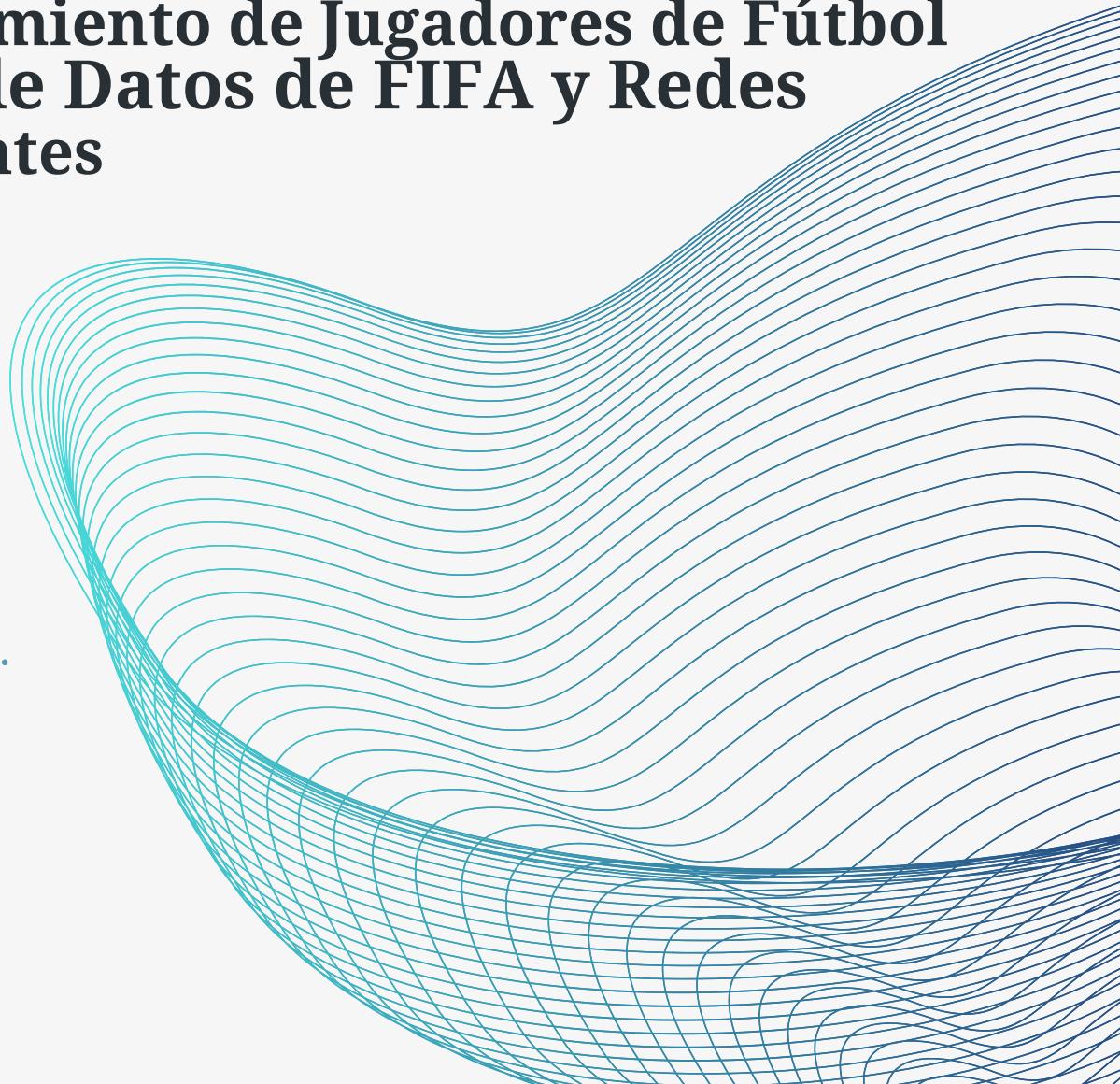




Predicción del Rendimiento de Jugadores de Fútbol Utilizando la Base de Datos de FIFA y Redes Neuronales Recurrentes

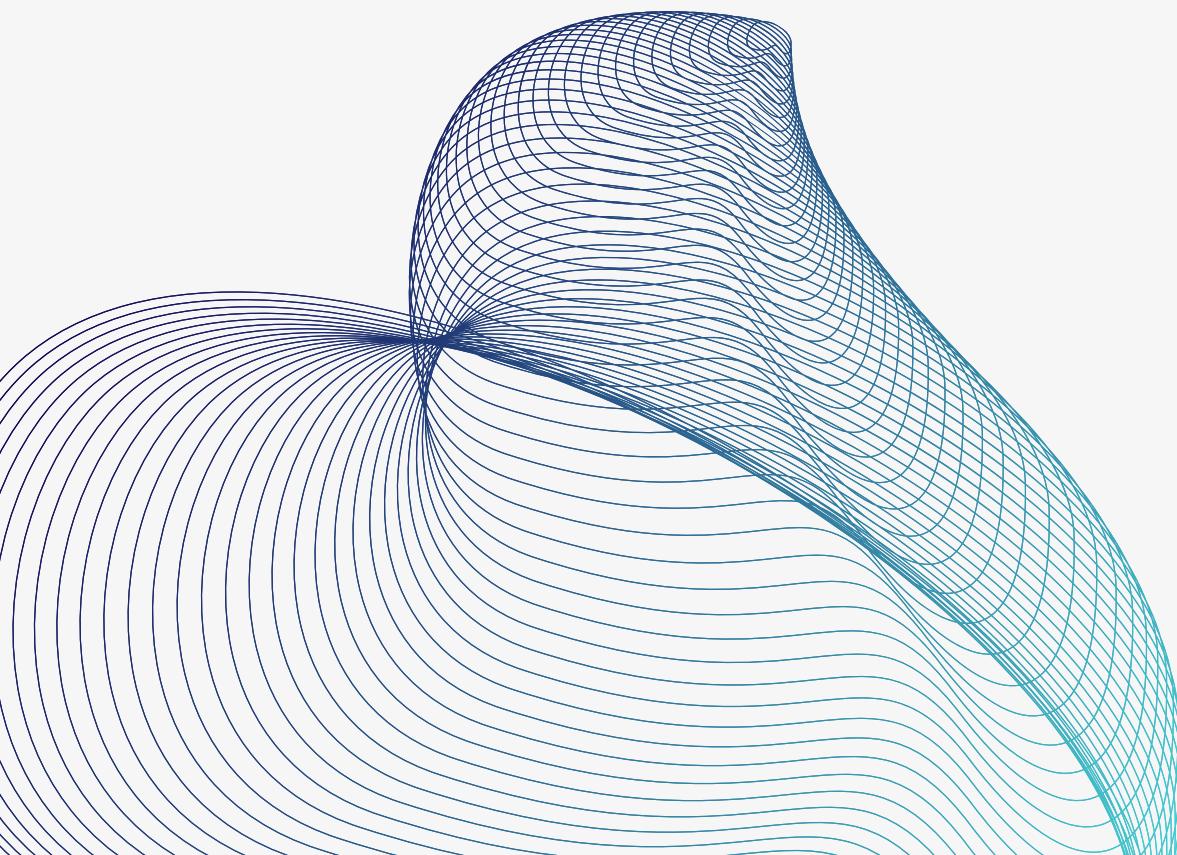
Juan Sebastian Rodriguez.
Alejandro Vega.





Modelo de RNN

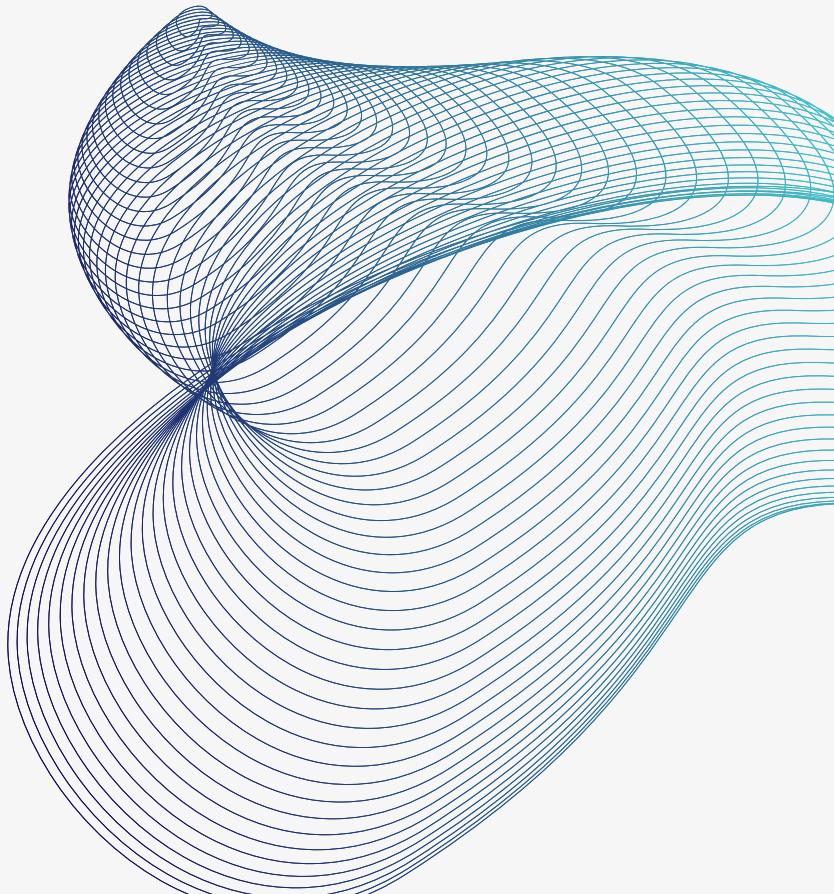
Desarrollar un modelo de RNN para predecir calificaciones futuras de jugadores, basado en atributos históricos.





Conjunto de Datos

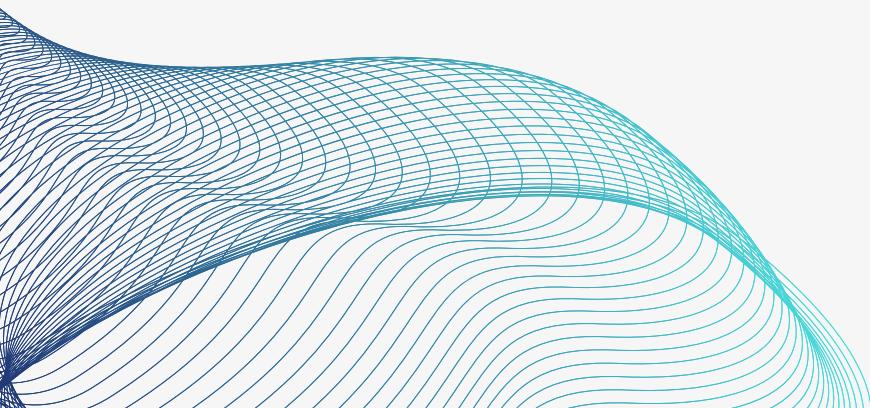
Base de datos “FIFA 2021 to
2005 Complete Player
Attributes” extraída de
Kaggle





Preprocesamiento de Datos

- Concatenación de archivos .csv de múltiples años y limpieza inicial con scripts en Python.
- Eliminación de valores nulos y duplicados; exclusión de los porteros para evitar ruido en el modelo.
- Estandarización de variables para asegurar una escala uniforme y mejora en el rendimiento del modelo.





Selección de Variables

Se toma current_rating como la variable objetivo y se toman las variables a usar en el modelo las que tengan 25% o más de correlación con la objetivo.

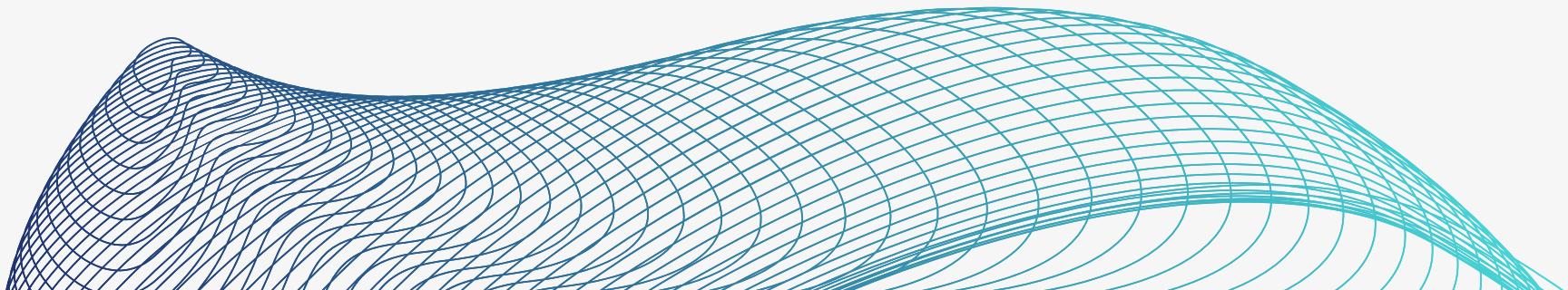
TABLE II
VARIABLES SELECCIONADAS PARA EL MODELO DE PREDICCIÓN

Variable	Descripción	Categoría
id	Identificador único del jugador	Identificación
Fullname	Nombre completo del jugador	Identificación
year	Año de la calificación	Temporal
current_rating	Calificación actual del jugador	Objetivo
potential_rating	Calificación potencial	Rendimiento
ball_control	Control del balón	Técnica
dribbling	Dribbling o regate	Técnica
marking	Marcaje defensivo	Defensa
aggression	Nivel de agresividad	Comportamiento
composure	Nivel de compostura	Comportamiento
reaction	Velocidad de reacción	Técnica
crossing	Habilidad para centros	Técnica
short_pass	Habilidad para pases cortos	Técnica
long_pass	Habilidad para pases largos	Técnica
acceleration	Aceleración	Físico
stamina	Resistencia	Físico
strength	Fuerza	Físico
spring_speed	Velocidad máxima	Físico
heading	Habilidad de remate de cabeza	Técnica
shot_power	Potencia de disparo	Técnica
finishing	Definición o finalización	Técnica
long_shots	Precisión en disparos lejanos	Técnica
FK_Acc.	Precisión en tiros libres	Técnica
reactions	Capacidad de reacción	Técnica
att_position	Posicionamiento en ataque	Técnica
interceptions	Habilidad para interceptar	Defensa
vision	Visión de juego	Técnica
jumping	Habilidad de salto	Físico
curve	Curvatura en el disparo	Técnica
fk_acc	Precisión en tiros libres	Técnica
penalties	Habilidad para penales	Técnica
age	Edad del jugador	Demográfico



Implementación de Secuencias Temporales

- Construcción de secuencias de cinco años para capturar tendencias a corto y mediano plazo.
- Cada secuencia representa los cambios en el rendimiento, ayudando al modelo a entender patrones históricos.





Modelos Iniciales de Machine Learning

Se utilizaron múltiples modelos de regresión como prueba inicial.

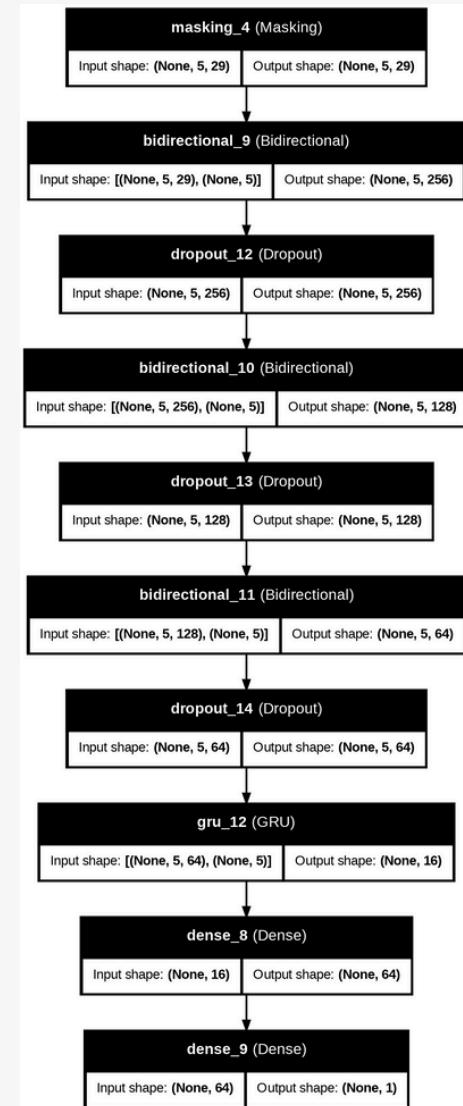
COMPARACIÓN DE MODELOS DE MACHINE LEARNING EN EL CONJUNTO DE PRUEBA COMPLETO

Modelo	MSE	R ²
Ridge Regression	3.846269×10^{-12}	1.000000
Lasso Regression	1.663278×10^{-4}	0.999997
Bayesian Ridge	1.952025×10^{-27}	1.000000
SVR	1.123554×10^{-3}	0.999998
Random Forest	1.512559×10^{-4}	0.999998
Decision Tree	0.1208985	0.998031
Regresión Polinomial (grado 2)	1.854907×10^{-9}	1.000000



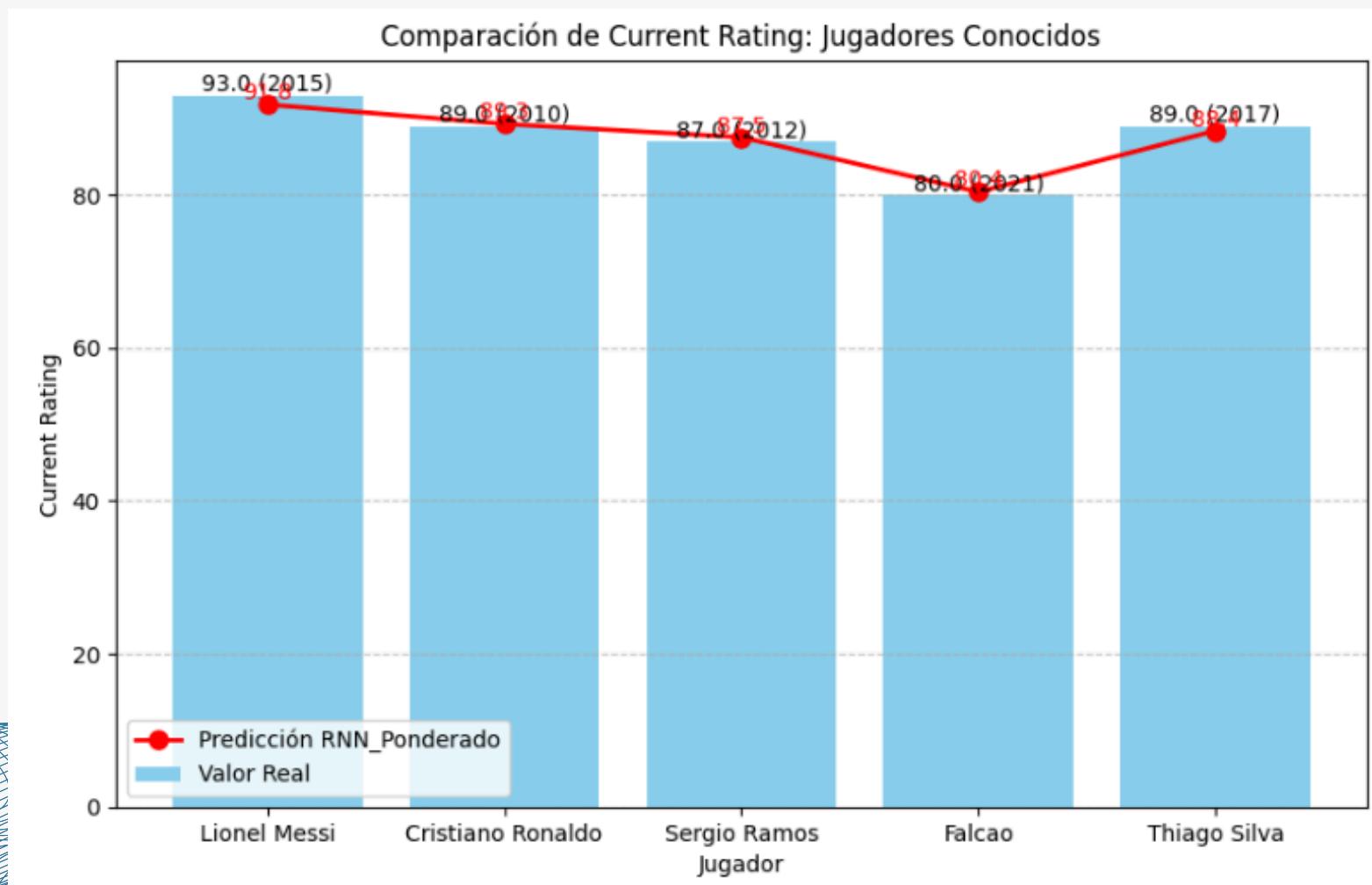
Desarrollo del Modelo RNN

- Implementación de un modelo RNN para capturar relaciones temporales en los datos.
- Mejora del modelo con una arquitectura RNN_Ponderado, que incluye capas LSTM y GRU bidireccionales.
- Ponderación de muestras para mejorar predicciones en jugadores con calificaciones altas.



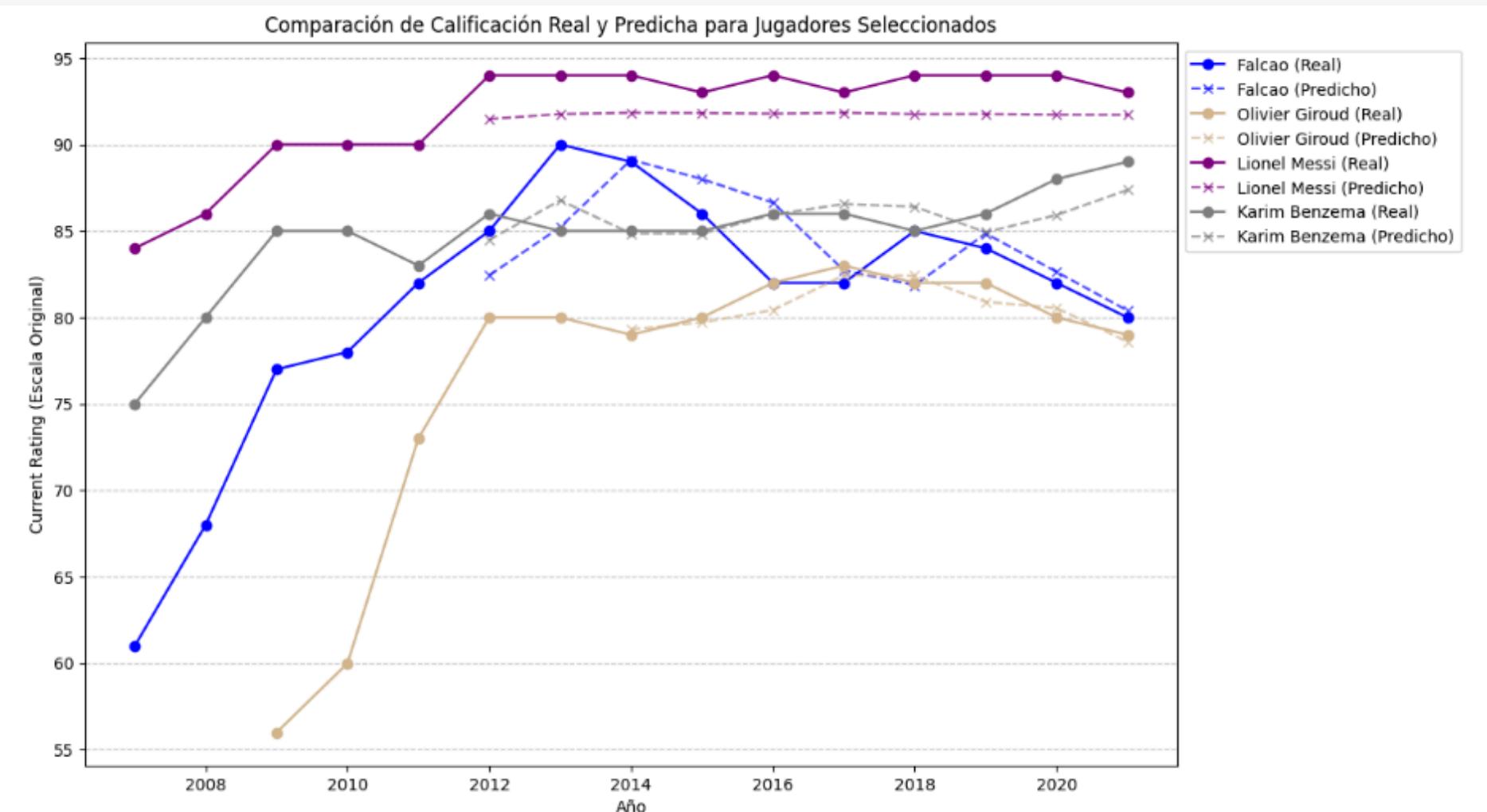


Resultados del Modelo RNN Ponderado





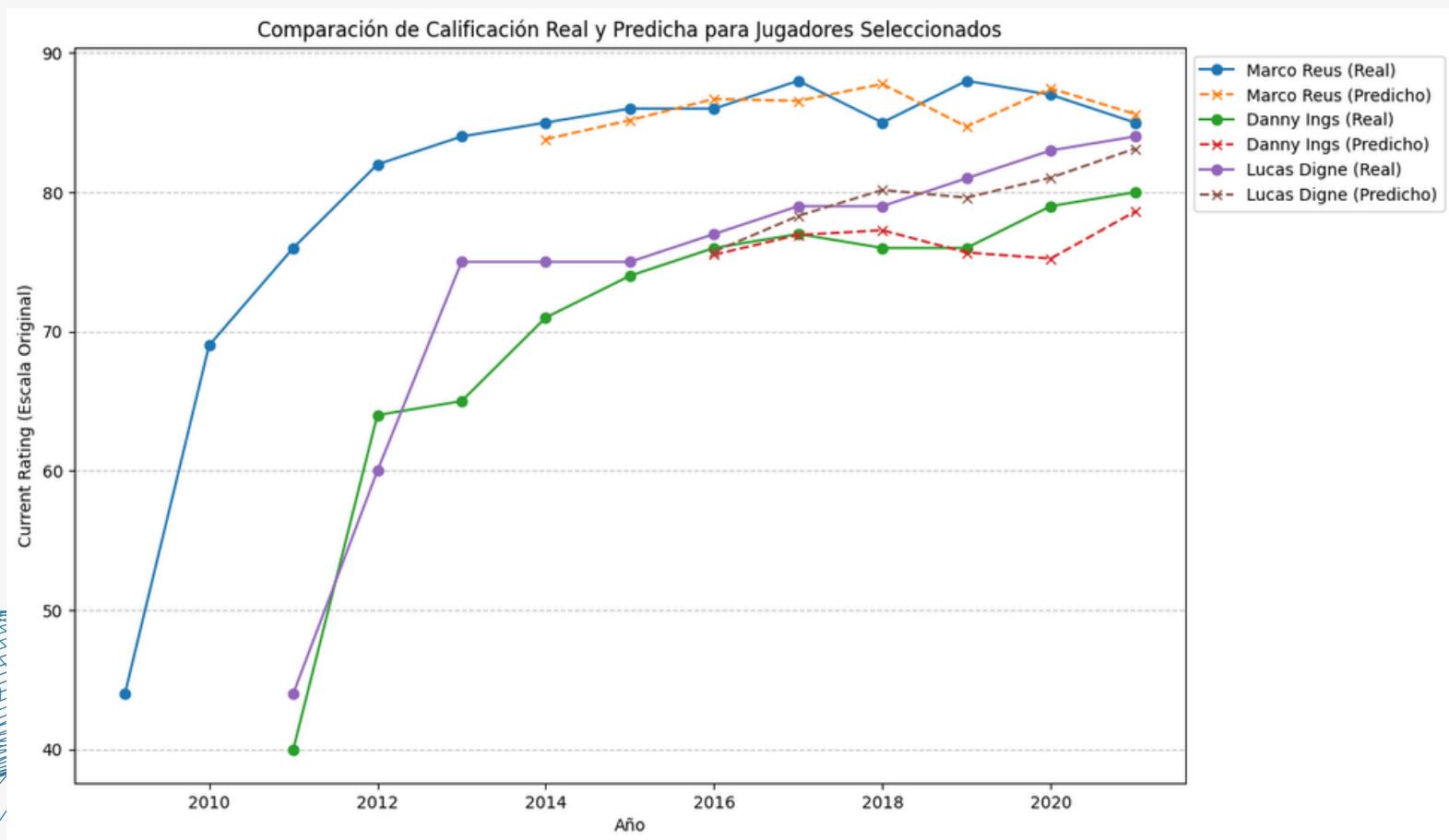
Resultados del Modelo RNN Ponderado





Análisis de Progresión y Proyecciones Futuras

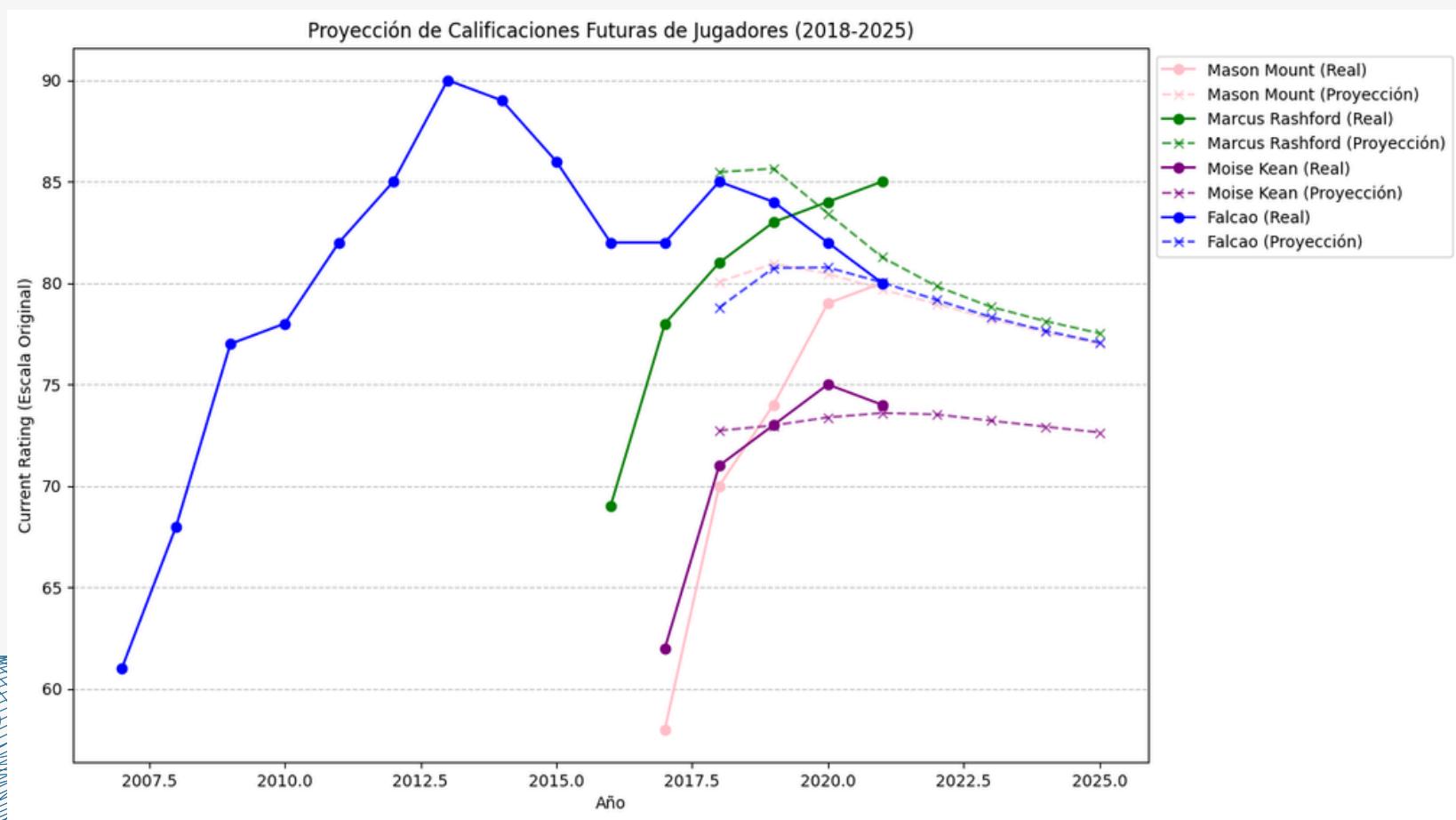
Análisis de jugadores con mayor progresión en su rendimiento.





Análisis de Progresión y Proyecciones Futuras

Proyecciones futuras para el período 2018-2025, mostrando capacidad del modelo para anticipar cambios en rendimiento.





Evaluación de Errores

- Se calcula el error absoluto medio por año y por posición.



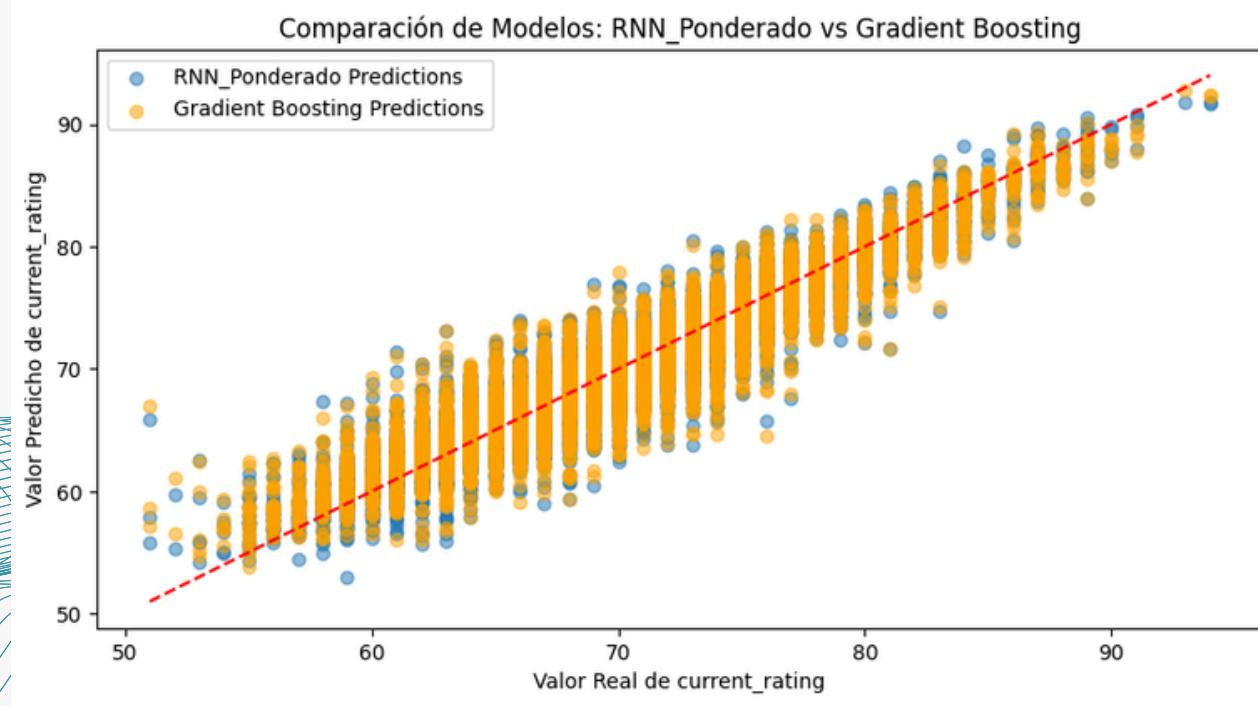
MAE POR AÑO

Año	MAE
2005	1.38
2007	1.39
2008	1.44
2009	1.32
2010	1.39
2011	1.56
2012	1.38
2013	1.49
2014	1.27
2015	1.32
2016	1.63
2017	1.20
2018	1.20
2019	1.91
2020	1.55
2021	1.52



Conclusiones y Trabajos Futuros

- Conclusiones: Buen desempeño del modelo, mejor en calificaciones altas. Limitación en predicciones para jugadores en desarrollo.
- Trabajos Futuros:
 - Usar datos reales de rendimiento.
 - Optimizar la arquitectura con modelos avanzados.
 - Probar el modelo en contextos competitivos y ampliar variables.





Muchas Gracias

