

PREDICCIÓN DEL TALENTO FUTBOLÍSTICO CON MACHINE LEARNING



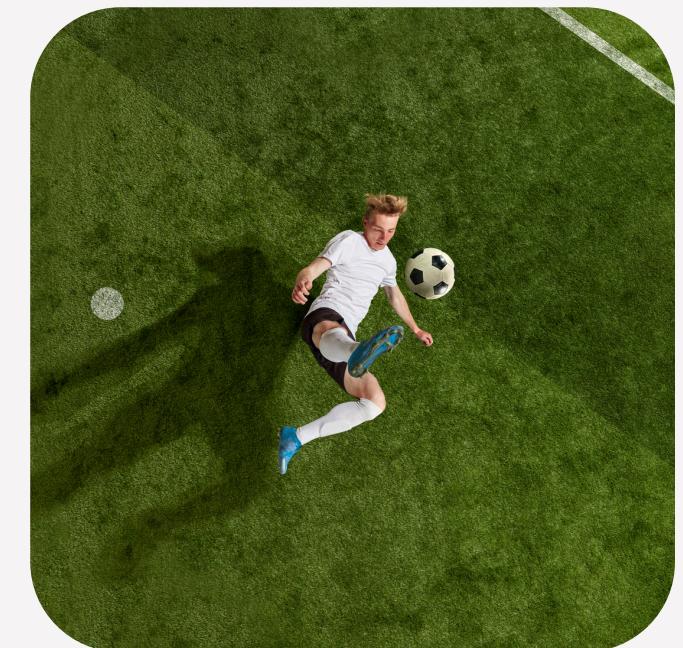
.....



20 DE MAYO 2025



.....



• • • •

INTRODUCCIÓN

LA GESTION DEL FUTBOL
IMPLICA GESTIONAR AMBOS
ASPECTOS DEPORTIVOS Y
EMPRESARIALES

KEY AREAS,

JUEGO DEL EQUIPO

SOSTENIBILIDAD FINANCIERA

MARKETING

VINCULACION DE
FANS

• • • •

EL GOL ES UN
TERMINO DE EXITO
DE LARGO
RECORRIDO
DENTRO Y FUERA
DEL TERRENO DE
JUEGO



PREDICCION DEL TALENTO PREDECIR EL TALENTO FUTBOLÍSTICO ES UN DESAFÍO APASIONANTE QUE COMBINA ANALISIS TECNICO, INTELIGENCIA TÁCTICA Y EVALUACION PSICOLOGICA.



.....



**ANÁLISIS DE
RENDIMIENTO:**

**FACTORES
PSICOLÓGICOS:**

**MÉTODOS DE
SCOUTING
TRADICIONALES:**



**EVALUACIÓN DE
INTELIGENCIA TÁCTICA:**

**USO DE TECNOLOGÍA
AVANZADA:**

HERRAMIENTAS COMO EL
ANÁLISIS DE VIDEO,
SENSORES DE RENDIMIENTO Y
SIMULACIONES DE REALIDAD
VIRTUAL AYUDAN A DETECTAR
PATRONES Y MEJORAR
HABILIDADES

.....

EL OBJETO DEL ESTUDIO

- ES PREDECIR LA VALORACIÓN OVERALL DE UN JUGADOR A PARTIR DE UNAS CARACTERÍSTICAS FÍSICAS Y TÉCNICAS.

LOS OBJETIVOS Y ALCANCE DEL PROYECTO

- SE EXTIENDEN A LA DETERMINACIÓN DE LA VARIABLE OVERALL A PARTIR DE ALGUNAS DE LAS VARIABLES DEL DATASET

• • • •

EL PROYECTO

EL PROPÓSITO DEL ANÁLISIS ES DEMOSTRAR O NO LA VIABILIDAD DE UTILIZAR DATOS DE LOS JUEGOS FIFA PARA LA PREDICCIÓN DE TALENTO FUTBOLÍSTICO MEDIANTE SU COEFICIENTE OVERALL



.....

EL DATASET



180021 FILAS

109 COLUMNAS

CSV

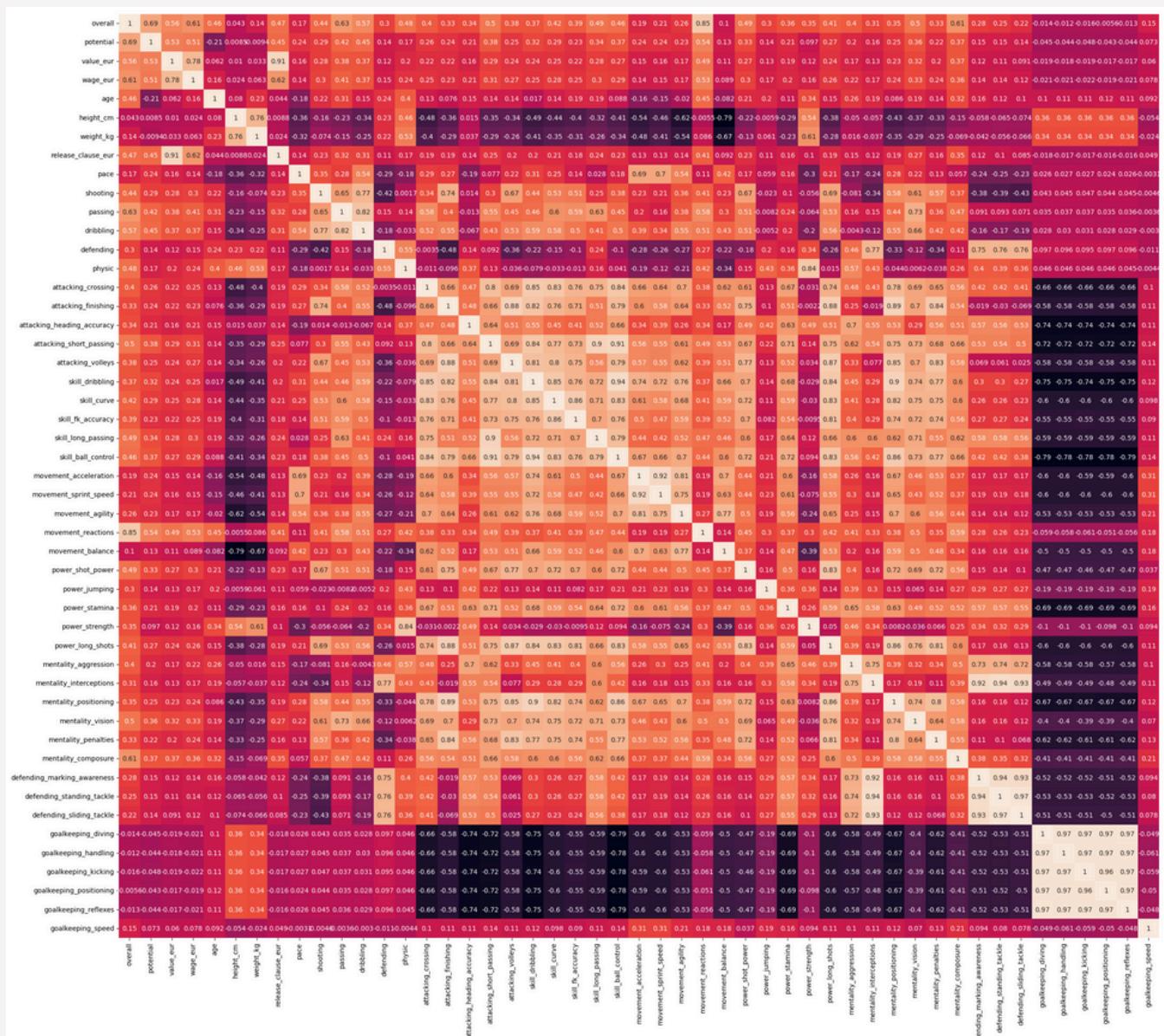
ESTADISTICAS DE
JUGADORES 2015-2024

EDA

DISTRIBUCIONES DE VARIABLES: SE DETERMINAN LAS VARIABLES MAS IMPORTANTES COMO EL VALOR OVERALL(GENERAL) Y EL POTENCIAL DE CADA JUGADOR, ADEMÁS DE LA EDAD, VALOR DE MERCADO Y SU SALARIO.

MEDIDAS ESTADÍSTICAS: PARA LAS COMPARATIVAS SE USA LA MEDIANA YA QUE ES LA MEDIDA QUE MEJOR SE AJUSTA A TODAS LAS VARIABLES

IDENTIFICACIÓN DE OUTLIERS: AL REALIZAR EL ANÁLISIS UNIVARIADO SE DETECTAN OUTLIERS EN LOS DATOS SOBRETODA EN LA VARIABLE VALOR DE MERCADO Y SALARIOS.



CORRELACIONES: SE ESTUDIAN LAS MISMAS



PREPROCESAMIENTO DE LOS DATOS

INTEGRER,FLOAT Y OBJECT DE FORMA CORRECTA A SU VEZ CONTAMOS CON VALORES FALTANTES PARA ALGUNAS VARIABLES COMO VALUE_EUR Y WAGE_EUR

NO SE DETECTAN REGISTROS DUPLICADOS QUE PUEDAN DISTORSIONAR EL MODELO



'POTENTIAL','VALUE_EUR','WAGE_EUR','AGE', 'PACE',
'SHOOTING', 'PASSING',
'DRIBBLING', 'DEFENDING'
Y 'PHYSIC'

TRATAMOS LOS FALTANTES APLICANDO LA MEDIANA POR CONSIDERARLA LA MAS CONVENIENTE.



MODELADO

MODELOS
SUPERVISADOS

**8 MODELOS UN MODELO
DE REGRESIÓN LINEAL,
UN RIDGE REGRESSION,
UN RAMDONFOREST, UN
XGBOOST, UN GRADIENT
BOOSTING, UN SVR, UN
LIGHTGBM Y UN
GRIDSEARCH DE
RANDOMFOREST
MIDIENDO SU MAE**

Modelo	Error de predicción	Tiempo de ejecución
Regresión Lineal	1.8178	55 sg
Ridge Regression	1.7403	1 sg
RandomForest	0.4339	6 min 54 sg
XGBoost	0.6148	0.4 sg
Gradient Boosting	0.6358	1 min 6 sg
SVR (kernel='rbf', C=100, gamma=0.1)	128 min	STOP
LightGBM	0.6405	0.6 sg
RandomForest con GridSearch	0.4434	5 min 32 sg





NO SE OBTIENEN
MEJORAS
SIGNIFICATIVAS

PRUEBAS Y MEJORA DEL MODELO

OVERFITTING

EL MODELO PRESENTA UN MAE DE 0.4376 PERO PRESENTA UN MAE DE ENTRENAMIENTO DE 0.1677 LO CUAL NOS LLEVA A INTERPRETAR QUE EXISTE OVERFITTING, EL MAE EN ENTRENAMIENTO ES BAJO (0.1677), LO QUE SUGIERE QUE EL MODELO APRENDE BIEN EN LOS DATOS DE ENTRENAMIENTO

ESTRATEGIAS

REDUCIR LA COMPLEJIDAD DEL MODELO BAJANDO MAX_DEPTH EN EL GRADIENT BOOSTING PARA EVITAR ÁRBOLES DEMASIADO PROFUNDOS



TAMBIÉN USAMOS EL CROSS_VAL_SCORE() PARA VERIFICAR SI EL MODELO ES ESTABLE EN DIFERENTES PARTICIONES DE DATOS Y CON UN VALOR DE 0.4508 PARECE QUE ES ASÍ



Modelo	Error de predicción	Tiempo de ejecución
RandomForest	0.4376	1 min 6 sg
XGBoost	0.6148	0.4 sg
Gradient Boosting	0.6358	1 min 6 sg
LightGBM	0.6405	0.5 sg
RandomForest con GridSearch	0.6765	1 min 8 sg

MODELO FINAL

RANDOMFOREST

MAE 0.4376

1 MINUTO 6SG

VARIABLES

'POTENTIAL','VALUE_EUR','WAGE_EUR','AGE', 'PACE', 'SHOOTING', 'PASSING', 'DRIBBLING', 'DEFENDING', 'PHYSIC'

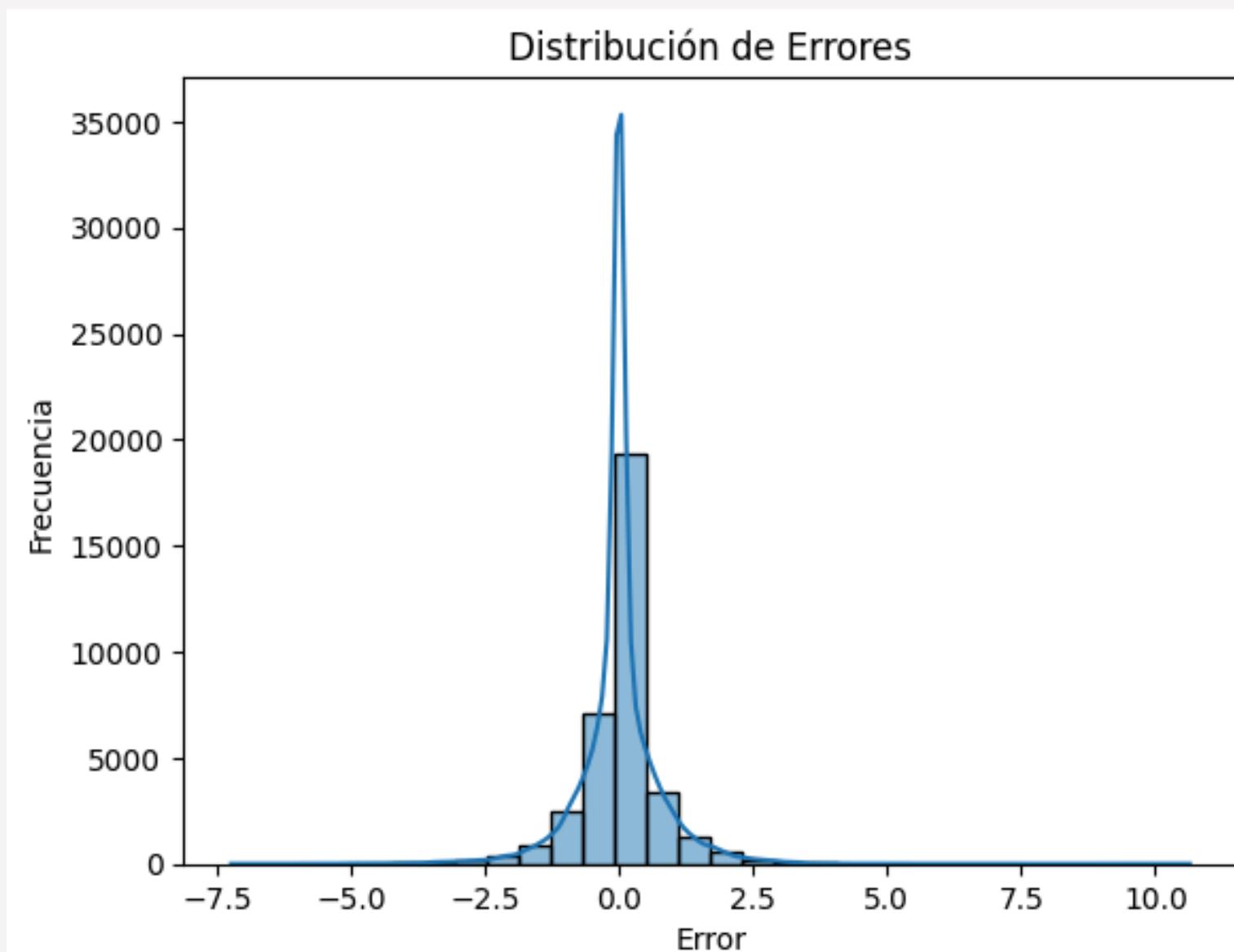
PUEDE CAPTURAR RELACIONES COMPLEJAS QUE LA REGRESIÓN LINEAL NO DETECTA

PROMEDIA EL RESULTADO DE MÚLTIPLES ÁRBOLES, LO QUE REDUCE EL RIESGO DE APRENDER PATRONES ESPURIOS.

PERMITE EVALUAR QUÉ CARACTERÍSTICAS TIENEN MÁS PESO EN LA PREDICCIÓN.

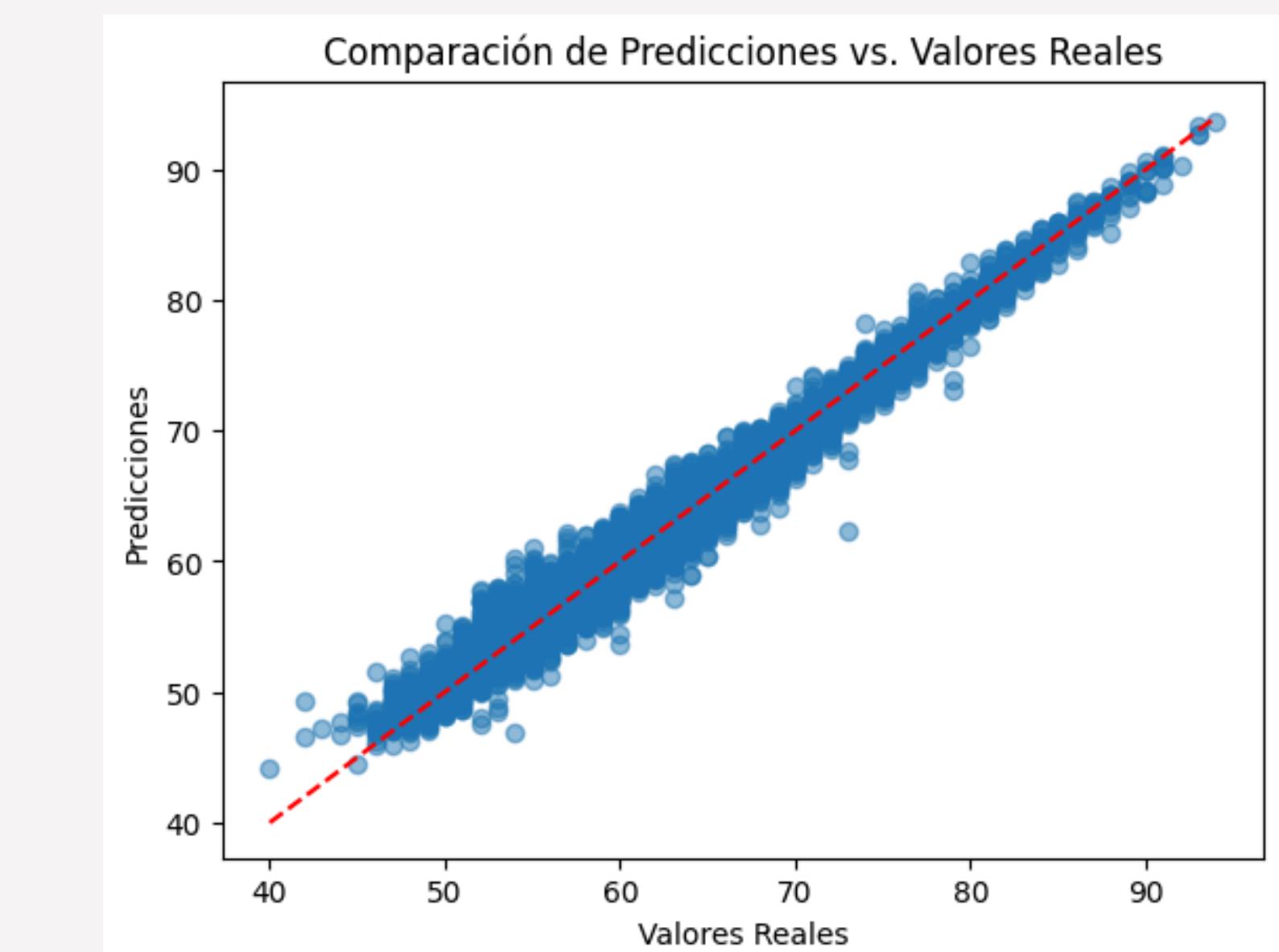
ESCALABILIDAD: FUNCIONA BIEN CON GRANDES VOLÚMENES DE DATOS SIN PERDER EFICIENCIA

DISTRIBUCION DE ERRORES



PREDICCIÓN Y RESULTADOS FINALES

PREDICCIÓN VS DATOS REALES





IMPACTO EN EL NEGOCIO

VENTAJA
COMPETITIVA

BENEFICIOS
ESTRATEGICOS

MEJOR TOMA DE
DECISIONES

REDUCCIÓN DE
COSTOS Y
OPTIMIZACIÓN DE
RECURSOS

FORTALEZAS VS DEBILIDADES



ALTA PRECISION Y
ESTABILIDAD

CAPACIDAD DE
MANEJAR DATOS CON
MUCHAS VARIABLES

ESCALABILIDAD

FLEXIBILIDAD EN LA
INTERPRETACION

DEMANDA
COMPUTACIONAL

MENOR
INTERPRETABILIDAD

RIESGO DE SOBREAJUSTE EN
CIERTOS ESCENARIOS

POSIBLE SENSIBILIDAD A DATOS RUIDOSOS

CONCLUSIONES Y PASOS FUTUROS

EL MODELO OFRECE UNA SOLUCIÓN SÓLIDA PARA LA PREDICCIÓN DE VALORES CON UN BUEN EQUILIBRIO ENTRE PRECISIÓN Y ROBUSTEZ

MARGEN DE MEJORA EN INTERPRETABILIDAD Y OPTIMIZACIÓN

AJUSTANDO
HIPERPARÁMETROS

USO DE MODELOS MÁS
INTERPRETABLES SI LA
EXPLICABILIDAD ES
CLAVE

MEJOR SELECCIÓN DE
CARACTERÍSTICAS
PARA REDUCIR RUIDO
EN LOS DATOS

