

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

“ANÁLISIS RETROSPECTIVO DEL RENDIMIENTO DE UN EQUIPO DE FUTBOL
PROFESIONAL COSTARRICENSE A PARTIR DE VARIABLES RECOLECTADAS
POR MEDIO DE TECNOLOGÍA INERCIAL”

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado en
Ciencias del Movimiento Humano y la Recreación para optar al grado y título de Maestría
Académica en Ciencias del Movimiento Humano.

PIER LUIGI MORERA SIÉRCOVICH

Ciudad Universitaria Rodrigo Facio, Costa Rica

2023

Dedicatoria

A mis padres Carlos Luis Morera Castillo y Flavia Siércovich Bartoli, quienes me inculcaron el valor del estudio y la importancia de perseverar para alcanzar las metas con pasión, disciplina y esfuerzo.

A mi esposa Karla Alvarado Rojas y a mis hijos, Gian Mauro y Alessandro, porque a ellos me debo! Todos mis esfuerzos y batallas tendrán siempre un motivo: su bienestar y felicidad.

P.L.M.S.

Agradecimientos

De forma especial quisiera agradecer al profesor tutor Ph.D. José Moncada Jiménez por brindarme su guía, su conocimiento, su paciencia y consejos en esta etapa final de la maestría.

Al Deportivo Saprissa, sus dirigentes, jugadores y compañeros del cuerpo técnico que me ofrecieron su colaboración para la obtención de la información.

También dar gracias a todas las personas que contribuyeron en este proyecto en diferentes áreas:

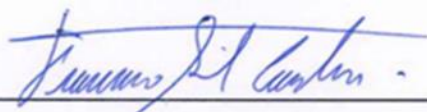
Exploración y análisis de datos: Dr. Francisco Siles Canales, M.Sc. Johnny Madrigal Pana, Bach. Daniely Hernández Orama, Bach. Nelson Torres Chávez, Ing. Daniel Méndez Zeledón, Bach. Steven Quirós Barrantes, Bach. Jairo Rodríguez Granados, Dr. Shu Wei Chou Chen, Bach. Silvia Campos Vargas, y Bach. Denia Vargas Araya.

Lectura, apreciación y comentarios: Ph.D. Pedro Carazo Vargas y Ph.D. Alejandro Salicetti Fonseca.

Correcciones filológicas: M.Sc. María Emilce Rojas Salazar (mi suegra).

Finalmente, agradecer a mi familia, por el apoyo incondicional y aliento continuo durante el proceso de investigación y redacción de esta tesis.

Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Ciencias del Movimiento Humano y la Recreación de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Ciencias del Movimiento Humano y la Recreación



Francisco Siles Canales, Dr.

**Representante del Decanato del
Sistema de Estudios de Posgrado**



José Moncada Jiménez, Ph.D.

Director de tesis



Pedro Carazo Vargas, Ph.D.

Asesor



Alejandro Salicetti Fonseca, Ph.D.

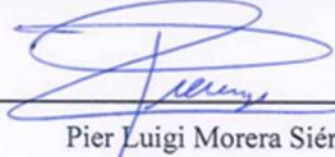
Asesor



Jessenia Hernández Elizondo, Ph.D.

Coordinadora

Programa de Posgrado en Ciencias del Movimiento Humano y la Recreación



Pier Luigi Morera Siércovich

Estudiante

Tabla de contenido

Dedicatoria	ii
Agradecimientos	iii
Tabla de contenido	v
Resumen	viii
Lista de tablas	x
Lista de abreviaturas	xiii
Capítulo I: Marco Teórico	1
1. Introducción	1
2. Análisis del rendimiento deportivo	3
2.1. Concepto de análisis del rendimiento deportivo.....	3
2.2. Evolución del análisis del rendimiento deportivo.....	4
2.3. Uso de técnicas y herramientas para la recolección de datos	5
2.3.1. La técnica de registro “manual”.....	5
2.3.2. Técnica de registro semiautomática	6
2.3.3. Técnica de registro automática	8
2.3.4. Desarrollo enfoques metodológicos en el análisis de rendimiento deportivo.....	11
3. Inteligencia artificial	13
3.1. Concepto de inteligencia artificial	13
3.2. Aprendizaje automático	14
3.2.1. Aprendizaje supervisado	14
3.2.1.1. Aprendizaje supervisado de clasificación.....	15
3.2.1.1.1. k vecinos más cercanos	15
3.2.1.1.2. Árboles de decisión	16
3.2.1.1.3. Bosques aleatorios.....	17
3.2.1.1.4. La máquina de soporte de vectores	19
3.2.1.2. Aprendizaje supervisado de regresión.....	20
3.2.1.2.1. Regresión logística	21
3.2.1.2.2. Regresión lineal múltiple.....	22
3.2.2. El aprendizaje no supervisado	22
3.2.2.1. Redes neuronales artificiales	23

3.2.2.2.	Análisis de componentes principales	23
4.	Análisis de la actividad del jugador de fútbol según su posición	25
4.1.	Variables técnico-tácticas.....	26
4.1.1.	Variables de ataque	27
4.1.1.1.	Gol/asistencia.....	27
4.1.1.2.	Pase	27
4.1.1.3.	Remate o Tiro al marco	28
4.1.2.	Variables defensivas	29
4.2.	Rendimiento físico.....	29
4.2.1.	Variables relacionadas con distancia	30
4.2.1.1.	Distancia total recorrida	30
4.2.1.2.	Distancia recorrida a diferentes intensidades.....	30
4.2.1.3.	Carrera de alta velocidad (HSR, “High speed running”).....	31
4.2.1.4.	Velocidad relacionada con Sprint	32
4.2.2.	Variables relacionadas con la carga	32
4.2.3.	Variables relacionadas con aceleración.....	33
4.3.	Análisis por posición	34
4.3.1.	Defensas	34
4.3.1.1.	Defensas centrales.....	34
4.3.1.2.	Defensas laterales.....	34
4.3.2.	Mediocampistas	35
4.3.3.	Delanteros.....	36
4.3.3.1.	Extremos delanteros	36
5.	Análisis de rendimiento en el fútbol profesional de Costa Rica	37
Capítulo II.	Justificación	40
Capítulo III.	Propósito.....	42
a.	<i>Objetivo General</i>	42
b.	<i>Objetivos Específicos</i>	42
Capítulo IV.	Metodología	43
a.	Diseño.....	43
b.	Participantes	43
c.	Instrumentos de medición	45
d.	Procedimientos	46
e.	Análisis estadístico	47
1.	Reducción de datos	47

2. Modelos de análisis	47
3. Software de análisis estadístico	49
Capítulo V. Resultados	50
a. Estadística descriptiva	50
b. Modelos de regresión por posición en el campo de juego	50
c. Modelos de regresión logística por posición en el campo de juego	57
d. Modelos de árboles de decisiones	62
e. Modelos de bosques aleatorios.....	63
Capítulo VI. Discusión	67
Capítulo VII. Conclusión y Recomendaciones	99
Capítulo VIII. Referencias	104
Anexo 1. Variables del WiMu Pro	115
Anexo 2. Variables del INSTAT Pro	116
Anexo 3. Producción académica durante el programa de maestría	122

Resumen

Propósito: El propósito del estudio fue identificar las variables que aumentan las posibilidades de ganar un partido y las variables que explican el rendimiento de futbolistas de un equipo de fútbol profesional costarricense, según la posición en el terreno de juego.

Metodología: Se construyó un archivo de 1037 casos de jugadores con datos recopilados durante dos temporadas y se realizó un análisis con algoritmos de machine learning a partir del análisis de datos recolectados con registros semiautomáticos y tecnología inercial.

Resultados: Para explicar el rendimiento de los futbolistas se construyeron modelos de regresión lineal múltiple que tuvieron valores predictivos de 0.67 para defensas centrales, 0.56 para defensas laterales, 0.71 para mediocampistas defensivos, 0.82 para extremos, de 0.80 para mediocampistas creativos, y 0.82 para centro delanteros. Las variables con mayor peso en la estimación fueron de naturaleza técnico-tácticas: goles y asistencias (defensas centrales, delanteros extremos y centro delanteros), tiros a portería (defensas centrales, mediocampistas creativos, y centro delanteros), disputas defensivas ganadas (defensas centrales y laterales), porcentaje de efectividad de los pases y disputas por arriba ganadas (laterales, mediocampistas ofensivos y extremos delanteros). En cuanto a la predicción de resultado, el método de bosques aleatorios permitió obtener mayor precisión en la estimación de la clasificación ($AUC > 0.5$) comparado con árboles de decisión. Las variables que explican mejor cuando se ganan los partidos fueron la máxima velocidad (todas las posiciones excepto los defensas centrales), el número de sprints (defensas centrales, laterales, mediocampistas ofensivos y centro delanteros), pases (defensa central y mediocampista ofensivo), la distancia recorrida a alta intensidad (defensas centrales y mediocampistas ofensivos), y la cantidad de metros por minuto recorridos (mediocampista defensivos, extremo delanteros, y centro delanteros). Conclusión: Los modelos estudiados fueron capaces de asignar una calificación a los jugadores según su posición en el campo e identificar las variables más asociadas con ganar los juegos.

Palabras claves: análisis de partidos, monitoreo, fútbol, posición de juego, rendimiento físico.

Abstract

Purpose: The purpose of the study was to identify the variables that increase the chances of winning a game and the variables that explain the performance of soccer players from a Costa Rican professional soccer team, according to the position on the field. **Methods:** A file of 1037 player cases with data collected during two seasons was analyzed with machine learning algorithms based on data collected with semi-automatic records and inertial technology. **Results:** Multiple linear regression models explained the performance of the soccer players, which showed predictive values of 0.67 for central defenders, 0.56 for wing defenders, 0.71 for defensive midfielders, 0.82 for wingers, 0.80 for creative midfielders, and 0.82 for center forwards. The variables with the most significant weight in the estimation were technical-tactical: goals and assists (central defenders, extreme forwards, and center forwards), shots on goal (central defenders, creative midfielders, and center forwards), defensive disputes won (central defenders and full-backs), percentage of pass effectiveness and disputes over the top won (full-backs, attacking midfielders and forwards). Regarding the prediction of the result, the random forest method allowed for obtaining greater precision in the estimation of the classification ($AUC > 0.5$) compared to decision trees. The variables that best explain won games were maximum speed (all positions except central defenders), the number of sprints (central defenders, wings, attacking midfielders, and center forwards), passes (central defender and attacking midfielder), the distance covered at a high intensity (central defenders and attacking midfielders), and the number of meters per minute covered (defensive midfielders, wingers, and center forwards). **Conclusion:** The models studied assigned a rating to the players according to their position on the field and identified the variables most associated with winning games.

Key words: match analysis, monitoring, soccer, playing position, physical performance

Lista de tablas

	Página
Tabla 1. Estadística descriptiva de los jugadores por posición	50
Tabla 2. Variables del modelo de regresión lineal ajustado para defensas centrales	51
Tabla 3. Variables del modelo de regresión lineal ajustado para defensas laterales	52
Tabla 4. Variables del modelo de regresión lineal ajustado para mediocampistas defensivos	53
Tabla 5. Variables del modelo de regresión lineal ajustado para mediocampistas ofensivos	54
Tabla 6. Variables del modelo de regresión lineal ajustado para extremos delanteros	55
Tabla 7. Variables del modelo de regresión lineal ajustado para centros delanteros	56
Tabla 8. Variables significativas para el modelo de regresión logística para defensas centrales.	57
Tabla 9. Variables significativas para el modelo de regresión logística para defensas laterales.	58
Tabla 10. Variables significativas para el modelo de regresión logística para mediocampistas defensivos.	59
Tabla 11. Variables significativas para el modelo de regresión logística para mediocampistas ofensivos.	60
Tabla 12. Variables significativas para el modelo de regresión logística para extremos delanteros.	61
Tabla 13. Variables significativas para el modelo de regresión logística para centros delanteros.	62
Tabla 14. Indicadores del modelo de árboles de decisiones.	62
Tabla 15. Indicadores del modelo de bosques aleatorios.	63

Tabla 16. Síntesis de las variables que explican el rendimiento del futbolista en el juego según posición.	99
Tabla 17. Síntesis de las variables que aumentan las probabilidades de ganar un partido según la posición.	100

Lista de figuras

	Página
Figura 1. Mapa conceptual del marco teórico	2
Figura 2. Frecuencia de los métodos y número total de métodos de análisis de datos utilizados por período de 5 años.	12
Figura 3. Ejemplo de algoritmo kNN para un $k=3$	16
Figura 4. Ejemplo de estructura básica de un árbol de decisión	17
Figura 5. Ejemplo de estructura de bosques aleatorios	18
Figura 6. Esquema de un hiperplano	19
Figura 7. Uso de la función Kernel	20
Figura 8. Figura de la función Sigmoide	21
Figura 9. Formación de juego más frecuentemente utilizada de equipo analizado	44
Figura 10. Variables más importantes para los defensas centrales	64
Figura 11. Variables más importantes para los defensas laterales	64
Figura 12. Variables más importantes para los mediocampistas defensivos	65
Figura 13. Variables más importantes para los mediocampistas ofensivos	65
Figura 14. Variables más importantes para los extremos delanteros	66
Figura 15. Variables más importantes para los centros delanteros	66
Figura 16. Ejemplo del cálculo del índice InStat PRO para un defensa lateral	70

Lista de abreviaturas

3D: acrónimo de 3 dimensiones. Representación gráfica que emplea las tres dimensiones del espacio (alto, ancho y profundidad).

A.R.: Análisis de rendimiento.

Acc: aceleraciones.

ANNs: “*artificial neural network*”, red neuronal artificial.

CD: Centro delantero

D(m): distancia recorrida en metros.

DC: Defensa central

Decc: número de desaceleraciones.

DL: Defensa lateral

DT: “Decision tree”, árbol de decisión.

ED: Extremo delantero

FIFA: Federación Internacional de Fútbol Asociado.

GNSS: “Global navigation satellite system”, Sistema global de navegación por satélite.

GPS: “Global positioning system”, Sistema de posicionamiento global.

HMLd: Abreviatura de “High Metabolic Load Distance”, es la distancia en metros recorrida por un jugador cuando su potencia metabólica está por encima de 25.5 W/kg.

HSRd “*high-speed running distance*”, distancia recorrida a velocidad superior a 21km/h.

Hz: hercio o hertz. Representa un ciclo por cada segundo, entendiendo ciclo como la repetición de un suceso.

MD: Mediocampista defensivo

MEMs: “Micro-electrical mechanical systems”, Sistema microelectromecánico.

ML: “*Machine Learning*” o aprendizaje de máquinas.

MLB: Major League of Baseball.

MLS: Major League of Soccer.

MO: Mediocampista ofensivo

NBA: National Basketball Association.

NFL: National Football League.

OR: “*odds ratio*” razón de probabilidades

PCA: “*Principal component analysis*”, Análisis de componentes principales.

PLSCA: “*Partial Least Squares Correlation Analysis*”, Análisis de correlación de mínimos cuadrados parciales.

SVM: “*Support vector machine*”, Máquina de soporte de vectores.

WIMU: Wireless Inertial Measurement Units.

Capítulo I: Marco Teórico

1. Introducción

El deporte ha visto en los últimos años una fuerte influencia de la ciencia que le ha permitido desarrollar atletas o deportistas con mejores rendimientos, metodologías de trabajo más completas y organizaciones cada vez más eficientes. Particularmente, el fútbol, al ser una especialidad que acapara a muchas personas (practicantes o seguidores) a nivel mundial, evidencia un gran interés en la producción de información científica.

El desarrollo de la computación y la estadística han permitido obtener datos que son de interés en diferentes áreas. Por ejemplo, en el deporte, y en este caso el fútbol, los clubes y selecciones nacionales utilizan esta información para mejorar el rendimiento de sus jugadores, el estudio de sus rivales, la detección de talentos, o incluso a nivel de mercadeo para mejorar sus ganancias económicas. Los aficionados y la prensa utilizan los datos para analizar y debatir las actuaciones de los equipos o jugadores y los apostadores en cómo poder predecir un ganador antes de que comience un partido.

Esto evidentemente ha convertido al fútbol en un deporte realmente competitivo y táctico, generando un gran desafío para entrenadores y analistas deportivos que deben conocer la realidad del deporte, sus demandas y las variables que puedan cambiar el curso del juego.

La presente tesis tiene como objetivo principal determinar las variables que explican el rendimiento de un equipo masculino de fútbol costarricense durante la competencia, a partir del análisis de datos retrospectivos recolectados con tecnología inercial y datos técnico-tácticos referentes a estadísticas del partido según la posición del jugador.

El poder detectar e identificar variables directamente relacionadas con la posibilidad de ganar un juego, permitiría optimizar la colocación del jugador dentro del campo de juego y favorecería la mejora de su progresión y rendimiento. Al hacer uso de tecnología novedosa para la recolección de información y recurrir a técnicas de análisis como el modelado predictivo con el fin de explicar el rendimiento del equipo, será necesario desarrollar en este capítulo, los avances que ha tenido el análisis de rendimiento en el deporte, desde el desarrollo tecnológico hasta la forma en cómo se analizan los datos.

Por lo tanto, en el siguiente marco teórico se conceptualizará el término análisis de rendimiento, se conocerá la evolución en los estudios en función del desarrollo de las herramientas para la recolección de los datos, y el gran aporte que hoy en día ofrece la inteligencia artificial y la minería de datos en el deporte, con sus métodos y técnicas estadísticas para lograr extraer información valiosa. Por último, se mencionará evidencia científica de los hallazgos en los análisis de la actividad del jugador de fútbol durante la competición, según la posición desde la dimensión física hasta la dimensión técnico-táctica (Figura 1).

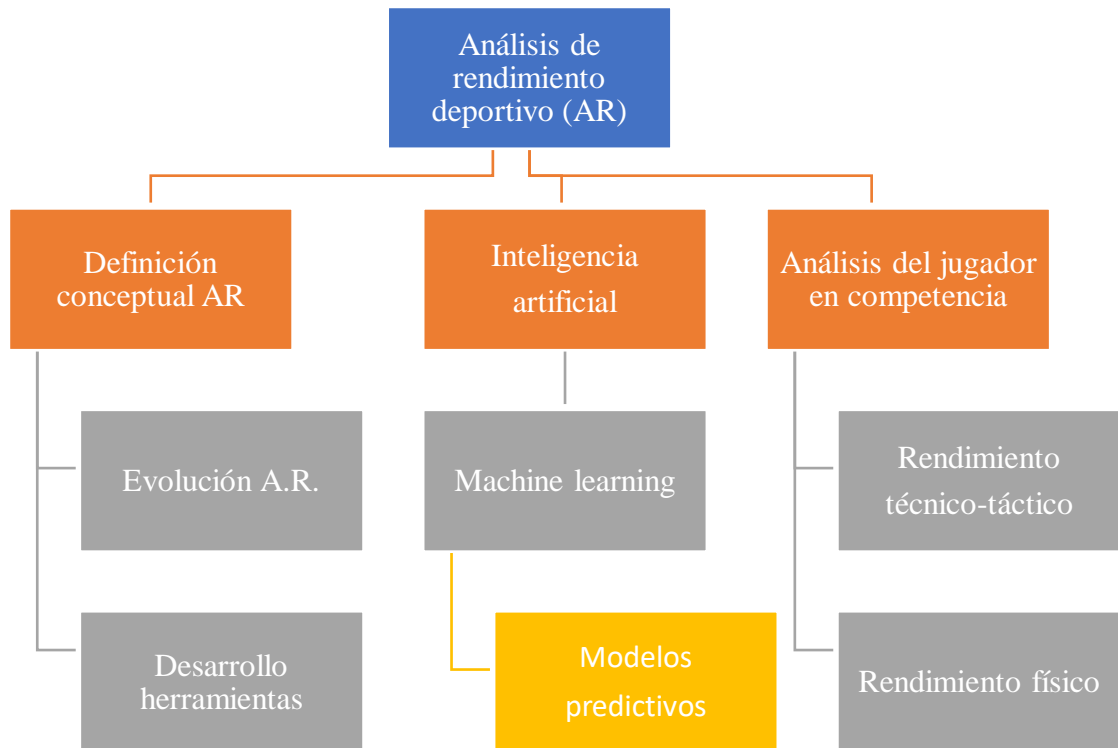


Figura 1. Mapa conceptual del marco teórico.

2. Análisis del rendimiento deportivo

2.1. Concepto de análisis del rendimiento deportivo

El análisis del rendimiento (AR) tiene su propio espacio en las Ciencias del Deporte, y se puede definir como “la investigación sobre la actuación deportiva en situación de competición y entrenamiento” (O'Donoghue, 2010; Rein & Memmert, 2016). En otras palabras, el estudio no se basa en situaciones aisladas de laboratorio, sino que busca con modelos de análisis mejorar los procesos de entrenamiento y el control de la actividad competitiva, prescribiendo el rendimiento de jugadores y equipos con base en las características de la propia competición (O'Donoghue, 2010; Rein & Memmert, 2016).

De esta manera, el AR tiene como objetivo identificar el desempeño, analizar los indicadores o variables claves del éxito (y el fracaso), además de presentar la información de una manera simple y concisa, de manera que el entrenador pueda implementar en la práctica estrategias efectivas de “coaching” para aumentar las posibilidades de éxito (Lord et al., 2020).

Esta comprensión del juego es fundamental para tomar decisiones más objetivas y de esta forma contribuir a la diferenciación entre las opiniones y los hechos. Generalmente, en el deporte, los juicios de desempeño son propensos al sesgo al no permitir analizar razonablemente, ya que los resultados buenos se atribuyen a causas internas; por ejemplo, darse el propio mérito cuando se obtiene la victoria y cuando los resultados negativos se adjudican a causas externas; es decir, negando la responsabilidad por los malos resultados, lo cual presumiblemente ocurre para mejorar o proteger la autoestima (Mark et al., 1984).

Se ha demostrado que los entrenadores son incapaces de recordar eventos de manera confiable durante un partido y que sólo pueden retener en la memoria un 30-40% de la información (Franks & Miller, 1991). Es un dato interesante que evidencia la necesidad por abarcar un área del estudio en el deporte que en los últimos años ha estado en pleno desarrollo. La comprensión de los indicadores de rendimiento no solo podría ayudar en el proceso de toma de decisiones de los entrenadores y jugadores con respecto al entrenamiento y la preparación del juego, sino también a otras partes interesadas relacionadas con el equipo, como por ejemplo la visoría (i.e., “scouting”) y a la propia administración, para evaluar correctamente el rendimiento del equipo y de los jugadores.

2.2. Evolución del análisis del rendimiento deportivo

El AR deportivo ha estado en los últimos años en crecimiento y constante desarrollo. Para conocer la realidad actual es necesario documentar la evolución que ha tenido a través del tiempo y así comprender los aprendizajes históricos e identificar los métodos más actuales.

Sarmiento et al. (2014), indican que se tiene documentado que el primer trabajo en esta área de investigación se llevó a cabo en 1910 por Hugh Fullerton en beisbol, y posteriormente la investigación científica fue mínima durante las siguientes décadas. Esta producción no sistemática y prolongada hasta principios de la década de 1990, pudo deberse a la escasez de manuales teóricos y metodológicos, así como la dispersión de la producción científica en revistas de carácter generalista.

En la década de los 90's ocurre un crecimiento fuerte en esta área, ya que se crearon sociedades científicas internacionales (e.g., Society of Performance Analysis of Sport), la edición de revistas científicas especializadas (e.g., Revista de Análisis del Rendimiento en el Deporte; Diario de Análisis Cuantitativo en el Deporte) y la constitución de departamentos de investigación en las unidades de educación superior, por lo que el AR pasó a ocupar un importante lugar en la producción científica en el área de las ciencias del deporte (Sarmiento et al., 2014). De hecho, en la última década, se ha comprobado un aumento sustancial en el número de publicaciones dirigidas en deportes colectivos (Lord et al., 2020) y una aplicación cada vez más fuerte de este conocimiento generado.

Hoy en día muchas instituciones deportivas han incorporado dentro de su organización un departamento de análisis de datos, o al menos una persona especializada en esta área. Isson (2018), citado por Moncada-Jiménez et al. (2021), menciona que, en Estados Unidos, la gran mayoría de equipos profesionales cuentan con esta figura. Por ejemplo, disciplinas deportivas como el fútbol americano (National Football League, NFL) un 56% de los equipos poseen su analista de rendimiento; en el baloncesto (National Basketball Association, NBA) cerca del 80%; y en el beisbol (Major League of Baseball, MLB) prácticamente todos sus equipos lo poseen (97%). En el fútbol, Moncada-Jiménez et al. (2021), refieren a la información publicada en el sitio web de la Major League of Soccer (MLS) (Major League of Soccer, 2019), donde destacan que aproximadamente 58% de los equipos poseen personal técnico especializado.

Dos aspectos pueden explicar muy bien la evolución que ha tenido esta ciencia. Primero, el uso de técnicas y herramientas sofisticadas para la recolecta de datos, y segundo, el enfoque metodológico que han utilizado los estudios.

A continuación, se hará un repaso de algunas técnicas e instrumentos novedosos que están siendo utilizados en el AR deportivo, ya que aportan una gran cantidad de información clave para comprender los fenómenos que ocurren durante el entrenamiento y durante la competición.

2.3. Uso de técnicas y herramientas para la recolección de datos

Las innovaciones tecnológicas han permitido una mayor disponibilidad de diferentes indicadores de desempeño. Un indicador de desempeño es una selección o combinación de variables de acción que tienen como objetivo definir algunos o todos los aspectos del desempeño, y pueden usarse para evaluar el desempeño de un individuo, un equipo o los elementos de un equipo (Geurkink et al., 2021). En la última década, los científicos han encontrado indicadores de rendimiento colectivo que obedecen a avances en las herramientas tecnológicas, a través de sistemas de seguimiento basados en análisis de video del movimiento de los atletas y la tecnología inercial.

Estos avances en las técnicas de registro, están muy bien explicados en el estudio de Castellano Paulis y Casamichana Gómez (2014a), quienes categorizan el proceso de codificación y registro de los datos según la intervención humana. De esta forma, ubican 3 grandes técnicas: a) registro manual, b) registro semiautomático, y c) registro automático.

2.3.1. La técnica de registro “manual”.

La técnica más común y sencilla se le denomina “papel y lápiz”. Los primeros estudios la utilizaban y consistía, por ejemplo, en ir registrando los desplazamientos y las acciones de los jugadores, empleando una hoja de anotación cuadrículada de un campo de fútbol con cuadrados a una escala milimétrica (escala 1:400) durante ciertos periodos de tiempo del partido. Las posiciones de los jugadores en el campo y las distancias recorridas por los mismos, se estimaban utilizando las medidas y marcas del campo, de manera que se sumaban los cuadros que había completado cada jugador (Castellano Paulis & Casamichana Gómez, 2014a). También se utilizaba para medir acciones básicas del juego como pases o tiros a puerta (Reep & Benjamin, 1968).

Más adelante, se combinó esta técnica con el uso de grabadoras de audio y consistía en verbalizar en una cinta magnetofónica los desplazamientos o movimientos que los jugadores realizaban. Reilly (1975), estimó distancias recorridas de acuerdo con la posición del jugador dentro del campo e identificaron cerca de 9000 actividades de movimiento durante un partido de fútbol con jugadores de la primera división inglesa. Además, lograron validar e idear un análisis de movimiento para la cuantificación de la tasa de trabajo.

Posteriormente, la computación ofreció una posibilidad aún más sencilla, ya que, a través de aplicaciones informáticas, se logra realizar observaciones en el lugar o a través del video, mejorando la calidad del dato registrado y aumentando el número de parámetros medidos. Inicialmente, esta técnica se llevaba a cabo mediante el teclado más común llamado QWERTY (i.e., el teclado común en las computadoras), en donde las teclas se programaban para que al pulsarlas registraran un código determinado, por lo que facilitaba la introducción de los datos y el proceso era aún más veloz. Luego apareció también el uso del “mouse”, y las tabletas digitalizadoras, que han ofrecido la posibilidad de ubicar las acciones en el espacio, aumentando así la precisión en el dato (Castellano Paulis & Casamichana Gómez, 2014a).

Esta técnica es de bajo costo, pero requiere de arduo trabajo (i.e., tiempo invertido en anotar) y puede propiciar errores humanos al introducir datos, pues se basa en las habilidades de los observadores que requieren un largo periodo de entrenamiento para obtener y analizar los datos de forma confiable (Hughes & Franks, 2004). Además, podría presentar algunos problemas para la codificación de cierto tipo de variables no directamente observables, como, por ejemplo, las velocidades o aceleraciones (Castellano Paulis & Casamichana Gómez, 2014a).

2.3.2. Técnica de registro semiautomática

La técnica de registro semiautomática utiliza el rastreo de imágenes de video (“video tracking”) para dar seguimiento a los desplazamientos de los jugadores por medio de un programa o software que permite hacerlo de manera semiautomática, y no necesita equipar al ejecutante con ningún tipo de instrumento (Bastida Castillo et al., 2018; Sarmiento et al., 2014). Esta técnica proporciona información de tiempo-movimiento (e.g. balón, jugadores y árbitros), y cuantifican de forma indirecta valores físicos (Bastida Castillo et al., 2018).

De esta manera, se han estudiado los desplazamientos de los jugadores profesionales durante el juego (e.g., caminar, trotar, correr, carrera a alta velocidad) (Dellal et al., 2011; Rampinini, Bishop, et al., 2007; Vigne et al., 2010) y también indicadores de rendimiento técnico-táctico (e.g., pases, tiros, expectativa de gol, posesión de balón en zona de ataque, estilos de juego) (Bush et al., 2015; Geurkink et al., 2021; Konefał, Chmura, et al., 2019a).

Dentro de los primeros estudios que utilizaron esta técnica, destaca el de Van Gool et al. (1988), citado por Castellano Paulis y Casamichana Gómez (2014a), realizado a finales de los años 80, en el que filmaron un partido amistoso de fútbol en película de cine de 16 mm con una frecuencia de grabación de 5 Hz (i.e., 5 imágenes por segundo) para digitalizar posteriormente los movimientos de todos los jugadores con el fin de encontrar la respuesta de la carga fisiológica demandada durante la competencia.

Hoy en día esta técnica de registro se ha comercializado y es posible acceder a la información de video tracking de jugadores, ligas y campeonatos a nivel mundial, a través de membresías o licencias. En los últimos años, han destacado las compañías Amisco® (francesa), ProZone® y Venatrack™ (inglesas), Tracab® (Suiza), Verusco® (Nueva Zelanda), Wyscout (Italia), e Instat (Rusia) (Arjol-Serrano et al., 2021; Redwood-Brown et al., 2012). Sin embargo, los reportes de los juegos no son inmediatos ya que se generan de 24 a 36 horas después de la finalización de los partidos (Castellano Paulis & Casamichana Gómez, 2014a).

Estos sistemas de medición semiautomáticos son sofisticados y de un elevado costo, ya que, en la mayoría de los casos, necesitan de una instalación fija de cámaras y de personal para el tratamiento de la información. Requieren la introducción de datos de forma “manual”, y debe verificarse continuamente si los jugadores son correctamente monitoreados por el programa informático e incluir las coordenadas cuando el sistema no haya sido capaz de hacerlo por sí mismo.

El sistema debe ser calibrado en términos de altura, longitud y anchura, para que el estadio y el campo sean transformados a un modelo bidimensional (2D) y así poder calcular la posición de los deportistas en un sistema de coordenadas (x, y) durante el partido, reconstruyendo los movimientos de los jugadores (Carling, 2001; Castellano Paulis & Casamichana Gómez, 2014a).

Al ser las imágenes de video una fuente de recopilación de información, la cantidad de luz puede limitar el registro, pero también algunas métricas no son posibles de valorar o medir ante la perspectiva o ubicación de la cámara (Bastida Castillo et al., 2018; Castellano Paulis & Casamichana Gómez, 2014a). Así, puede ocurrir un “bloqueo visual”, lo que hace que el sistema pierda el rastro de algunos movimientos. Por ejemplo, cuando los jugadores se juntan en una pequeña zona del campo al mismo tiempo, se “tapan” unos a otros lo cual “confunde” el sistema de rastreo. A este fenómeno se le llama oclusión, e implica la necesidad de agregar una etapa de corrección manual ya que las plataformas informáticas por sí mismas no logran resolverlo. Las oclusiones son problemas que deben resolver los programadores especializados en el rastreo de imágenes.

2.3.3. Técnica de registro automática

Actualmente, la digitalización es llevada a cabo en directo y en tiempo real, sin necesidad de personal que realice el registro de forma manual, automatizando el proceso, gracias a la innovación de sistemas micro tecnológicos (Bastida Castillo et al., 2018). La información se obtiene a través de pequeños dispositivos de poco peso, que son fáciles de transportar en la prenda, con una capacidad para medir y grabar una amplia cantidad de variables biométricas del deportista (Moncada-Jiménez et al., 2021). Estos datos se transfieren en tiempo real a través de una antena conectada a una computadora portátil o a una estación de trabajo.

Estos dispositivos utilizan la tecnología inercial ya que incorporan receptores de posicionamiento que son alimentados a través de sistemas locales (uso de antenas especiales) o del Sistema Global de Navegación por Satélite o GNSS (“Global Navigation Satellite System” por sus siglas en inglés).

El sistema de posicionamiento local o LPS (“*Local Position System*”, por sus siglas en inglés), dispone de un alto número de mediciones y exactitud de los datos recopilados en tiempo real. Además, cuenta con tecnología de banda ultra ancha que reduce interferencias en la vía de transmisión. Su desventaja principal es que necesita de una instalación fija de antenas que implica gastos económicos de instalación y de tiempo.

El GNSS es una red de satélites que transmite rangos de señales en cualquier parte del globo terrestre, que permiten a través del receptor, proporcionar información sobre la posición, la hora, la velocidad y la dirección a la que se mueve un objeto, las 24 horas del día

y en todas las condiciones climáticas (Castellano Paulis & Casamichana Gómez, 2014b). Está compuesto por dos sistemas de navegación diferentes: el Sistema de Posicionamiento Global (GPS, por sus siglas en inglés) de los Estados Unidos y el Sistema Orbital Mundial de Navegación por Satélite (GLONASS, por sus siglas en inglés) de la Federación Rusa.

La red GPS es más comúnmente utilizada ya que está disponible gratuitamente sin cargo ni restricción. El sistema fue diseñado por el departamento de defensa de los Estados Unidos con fines militares, y opera desde 1995. Este sistema está compuesto por 27 satélites en órbita a 11000 millas náuticas alrededor del planeta en seis trayectorias diferentes, cada una de las cuales realiza una revolución en 12 horas, dando una cobertura continua; es decir, hasta ocho satélites accesibles desde cualquier punto de la Tierra (Castellano Paulis & Casamichana Gómez, 2014b).

Sin embargo, pueden presentar algunas limitaciones que deben ser analizadas, y controladas, para garantizar una buena calidad del dato, por ejemplo, las interferencias en la señal de los satélites por estructuras altas (e.g., edificios, techos cercanos al terreno de juego), la refracción atmosférica, el número de satélites disponible para el receptor (mínimo cuatro satélites) y la geometría de disposición de los satélites entre sí y el receptor que afecta la calidad de triangulación de posición (Castellano Paulis & Casamichana Gómez, 2014b; Springham et al., 2020; Witte & Wilson, 2004). Ambos sistemas permiten cuantificar las variables de movimiento de los jugadores referente a la distancia total y relativa recorrida en diferentes zonas de velocidad, velocidad máxima y aceleraciones (Andrzejewski et al., 2018; Malone et al., 2018; Springham et al., 2020).

Los dispositivos inerciales también incorporan Sistemas Microelectromecánicos o MEMS, que cuentan con sensores inerciales como el acelerómetro, el giróscopo y el magnetómetro, que han sido utilizados recientemente con éxito en el deporte (Andrzejewski et al., 2018; Malone et al., 2018; Springham et al., 2020). Estos sensores permiten grabar valores sobre impactos y cargas, aceleraciones y cambios de dirección; los cuales son característicos de los deportes intermitentes como el fútbol (Castellano Paulis & Casamichana Gómez, 2014b; Jaspers et al., 2018).

Incluso es posible obtener otras mediciones como la altura, la frecuencia cardíaca (portando una banda torácica) y el costo metabólico o gasto energético de las acciones del

jugador, monitorizando así la demanda energética de la actividad desarrollada (Osgnach et al., 2010).

De esta manera, la cantidad de variables que se obtienen por atleta son muy grandes, incluso superiores a las 120 (Moncada-Jiménez et al., 2021). En cuanto a los informes de métricas derivadas de estos sistemas de seguimiento se pueden realizar de varias maneras (Torres-Ronda et al., 2022):

- a. Valores absolutos: describen métricas por todo el partido, o duración de la actividad deportiva (e.g., total de distancia recorrida durante el partido).
- b. Duraciones temporales: estratificando los partidos o entrenamientos por períodos basados en duraciones temporales (e.g., cada 5 o 10 min) para capturar la fluctuación en la carga externa a lo largo del partido o las características de máxima exigencia, llamadas “peor escenario posible”.
- c. Promedios segmentarios: que muestran el promedio de un valor durante un determinado tiempo (e.g., el enfoque de media móvil o “moving average”).
- d. Datos normalizados: Por ejemplo, clasificar por fases de juego: ataque y defensa o por posesión del balón.

Se mencionada que es tarea de los profesionales e investigadores considerar el más adecuado sistema de medición de acuerdo al deporte específico, el entorno y la aplicación prevista de los datos (Torres-Ronda et al., 2022).

El desarrollo de la técnica automática ha reportado en la literatura una gran confiabilidad y validez ya que permite una medición de patrones de movimiento muy propios de la actividad del fútbol como cambios de dirección o actividades de alta intensidad (Rampinini et al., 2015; Scott et al., 2016), aceleraciones (Akenhead et al., 2016), desaceleraciones y actividades de alta velocidad (Tierney et al., 2016), sprints de 10-30 m e incluso velocidades mayores a 30 km/h (Muyor et al., 2018), obteniendo muy buenos resultados.

La Federación Internacional de Fútbol Asociado (FIFA), en el año 2015 estableció dentro del reglamento de juego permitir el uso en partidos competitivos de los dispositivos de seguimiento inercial (Tierney et al., 2016), lo que ha generado un significativo aumento en la cantidad de investigaciones llevadas a cabo con esta tecnología. Esto se ha visto

reflejado en la base de datos PUBMED, que pasó de 3 a 136 artículos publicados por año entre 2001 y 2018 (Malone et al., 2020).

Por lo tanto, la implementación cada vez más desarrollada de las herramientas para el AR, brindan la posibilidad de encontrar variables innovadoras que permiten abrir nuevos horizontes. La facilidad con la que se recolectan los datos y la gran cantidad de información que generan estos dispositivos, no permitirían análisis estadísticos tradicionales, por lo que ha existido una necesidad de crear métodos o enfoques metodológicos de análisis más completos (García-Aliaga et al., 2021; Lord et al., 2020).

2.3.4. Desarrollo enfoques metodológicos en el análisis de rendimiento deportivo

En el año 2011, en una revisión de literatura publicada en la *Revista Portuguesa de Ciencias del Deporte*, Marcelino et al. (2011) muestran un análisis cronológico, categorizando los estudios y sistematizando las tendencias evolutivas, en tres distintos apartados. Primero, encontraron estudios predominantemente descriptivos, que concluían en datos globales de rendimiento, “sin mayor esfuerzo de contextualización” con el objetivo principal de identificar, describir y caracterizar patrones de movimiento, sobre todo en deportes de invasión como el fútbol, el hockey, el waterpolo y el rugby.

Posteriormente, aparecieron estudios de carácter comparativo, donde se analizaban distintas variables según la posición de los jugadores en el campo, el nivel competitivo, la diferencia de puntuación o la diferencia en el género. También se enfocaron en considerar variables situacionales con el fin de comprender los momentos críticos del juego, tomando en cuenta el tiempo del partido y la evolución del rendimiento durante el mismo, asociándolos con factores contextuales como la localización del partido, el tipo de competición, los oponentes (i.e., débiles o fuertes), o el tipo de partido, entre otros (Lago-Peñas et al., 2010; Lago-Peñas et al., 2011; Taylor et al., 2008). Esto representó un avance significativo ya que supuso una comprensión del juego y dio luz sobre la importancia del entrenamiento específico. Sin embargo, aún estos análisis no lograban explicar ni predecir el rendimiento deportivo.

Es en la última década que se inicia una fuerte tendencia en los estudios relacionados con la predicción de resultados, buscando un cambio de paradigma en cuanto a la forma de analizar y comprender los fenómenos deportivos, la variabilidad y la inestabilidad del juego,

considerando de forma más sistemática la influencia de las variables situacionales en el rendimiento deportivo (Lord et al., 2020; Marcelino et al., 2011; Sarmiento et al., 2014).

Este aumento significativo en los estudios de tipo predictivo, es evidenciado en el estudio de Lord et al. (2020), quienes analizaron la evolución de las técnicas de análisis de rendimiento a lo largo del tiempo (1997 a 2019), identificando los métodos utilizados y la frecuencia con la que fueron empleados (Figura 2). Los investigadores destacan los estudios que utilizan el análisis inferencial (i.e., diferencias entre el desempeño exitoso y el menos exitoso) como el método más utilizado en el área del AR deportivo y demuestran un predominio de estudios de tipo descriptivo en los primeros dos períodos observados (1997-2000 y 2001-2005).

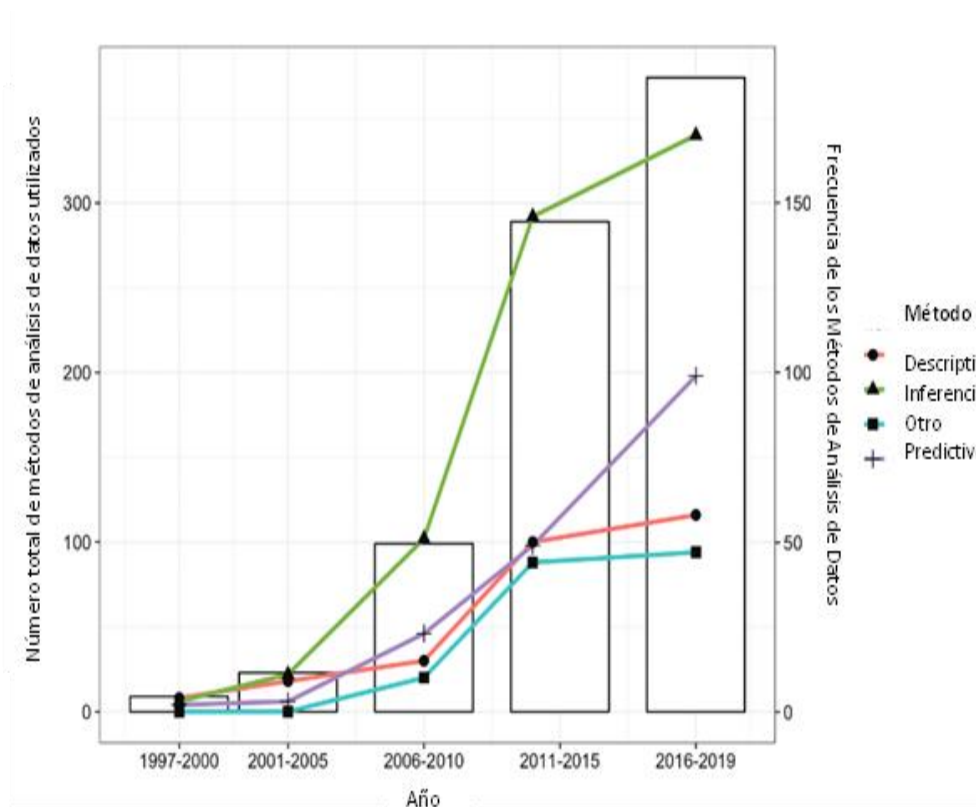


Figura 2. Frecuencia de los métodos de análisis de datos utilizados mostrados en líneas y el número total de métodos de análisis de datos mostrados en columnas por quinquenios (Lord et al., 2020).

Posteriormente, en los siguientes quinquenios (2006-2010, 2010-2015 y 2016-2019) se observó un aumento sustancial en la frecuencia de estudios que utilizaron el método predictivo, siendo aún más evidente en el período comprendido entre el 2016 y 2019. Esto

obedece a un cambio en la estrategia de descubrir conocimiento a través del aporte de una nueva herramienta dentro de la ciencia de datos, conocida como inteligencia artificial (García-Aliaga et al., 2021).

En el próximo capítulo se abordan algunas técnicas estadísticas y computacionales de la inteligencia artificial que han sido utilizadas en el deporte para el análisis de grandes volúmenes de datos.

3. Inteligencia artificial

3.1. Concepto de inteligencia artificial

De acuerdo con García-Aliaga et al. (2021), la inteligencia artificial “... es una expansión de la inteligencia humana que consiste en aprovechar las capacidad computacionales de las máquinas para apoyar a los humanos en sus actividades” (p. 49). Por su parte, Kaplan y Haenlein (2020) la definen como “la capacidad de un sistema para interpretar correctamente datos externos, aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible” (p. 40).

Es través de la inteligencia artificial que las computadoras aprenden, conforme se genera una mayor calidad y cantidad de datos, paralelamente se obtiene un aumento continuo y exponencial del poder de procesamiento, que para el ser humano es imposible realizar. Por esta razón ha tenido una atención considerable en la sociedad en general, debido a la inminente necesidad de transformar estos altos volúmenes de datos en conocimientos útiles.

La solución a estos altos volúmenes y combinación de varias fuentes de datos puede presentarse a través del reciente auge de las tecnologías de macrodatos o *Big data* (en inglés). El término *Big data*, aún no tiene una definición universalmente acordada y se describe más bien por las características que posee (Rein & Memmert, 2016): a) volumen, b) variedad, y c) velocidad. El volumen describe la magnitud de los datos (e.g., 400 gigabytes de datos de seguimiento en una sola temporada), la variedad se refiere a la heterogeneidad de los datos (e.g., posición, estado físico, entrenamiento, datos de video) y la velocidad, al patrón de procesamiento que caracteriza la tasa de producción de datos.

La minería de datos requiere de operaciones que deben ser analizadas por un estadístico, o quien conozca no solo los conceptos sino también sepa interpretar los datos cuando existen cambios (Bouza & Santiago, 2012). Por ello, se requiere de una interacción

entre las áreas de computación, estadística y del área donde se aplica a través de los expertos, para una correcta depuración de los datos y así lograr hacer una buena interpretación y contextualización de los resultados (Blei & Smyth, 2017).

Los profesionales han desarrollado métodos o enfoques estadísticos que detectan patrones en los datos y utilizan patrones descubiertos para predecir datos futuros o facilitar la toma de decisiones en entornos con incertidumbre; por ejemplo, el aprendizaje de máquinas o “machine learning” (ML) (Murphy, 2012).

Los métodos del aprendizaje automático han demostrado su eficacia en la medición de varios deportes colectivos (Claudino et al., 2019; Lord et al., 2020) ya que se han realizado una gran cantidad de estudios de predicción o modelado en la predicción de lesiones (López-Valenciano et al., 2020; Rossi et al., 2018), la predicción del rendimiento (Memmert et al., 2017; Tierney et al., 2016), la carga (Jaspers et al., 2018; Springham et al., 2020) o para caracterizar el efecto de una variable sobre la probabilidad de éxito (Cintia et al., 2015; Geurkink et al., 2021; Groll et al., 2018). Esto, evidentemente genera una ventaja competitiva, ya que ayuda a los profesionales que trabajan en la optimización del entrenamiento y en las estrategias de competición (Moncada-Jiménez et al., 2021).

3.2. Aprendizaje automático

El aprendizaje automático o ML es considerado un subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar sistemas computacionales diseñados con el propósito de aprender y adaptarse a partir de los datos, sin la necesidad de programar explícitamente el nuevo conocimiento adquirido.

De esta forma, se crean códigos para que la computadora comprenda qué es lo que se quiere transmitir. Estos códigos se generan a partir de algoritmos, que son un conjunto ordenado de operaciones sistemáticas que tratan, a partir de cálculos, encontrar respuesta a problemas. Existen dos modalidades de algoritmos: a) aprendizaje supervisado y b) aprendizaje no supervisado (e.g., agrupación) (Claudino et al., 2019).

3.2.1. Aprendizaje supervisado

En este tipo de técnica, se requieren datos de entrada y de salida para desarrollar un modelo predictivo, de forma tal que el algoritmo se entrena al otorgarle las preguntas (denominadas características) y las respuestas (denominadas etiquetas). Esto se hace con la finalidad de que el algoritmo las combine y pueda hacer predicciones. Algunas técnicas de

aprendizaje supervisado más populares en el análisis de rendimiento deportivo se describen a continuación.

3.2.1.1. Aprendizaje supervisado de clasificación

Una de las tareas comunes de aprendizaje automático, que implica predecir una variable de destino en datos nunca antes vistos, es la clasificación. El objetivo de la clasificación es predecir una variable objetivo (clase) mediante la construcción de un modelo de clasificación basado en un conjunto de datos de entrenamiento y luego utilizar ese modelo para predecir el valor de la clase de datos de prueba (Bunker & Thabtah, 2019). A continuación, se mencionarán algunas técnicas de clasificación.

3.2.1.1.1. k vecinos más cercanos

El algoritmo k vecinos más cercanos o kNN (“k-Nearest Neighbor” por sus siglas en inglés), es un algoritmo de clasificación más básico y esencial en el ML. Es un tipo de aprendizaje basado en instancias o aprendizaje perezoso, en el que la función sólo se aproxima localmente y todo el cálculo se difiere hasta la clasificación. Utiliza un método no paramétrico, o sea que no realiza suposiciones acerca de la distribución de los datos (Moncada-Jiménez et al., 2021). En un esquema común de ponderación consiste en dar a cada vecino un peso de $1/d$, donde d es la distancia al vecino, de modo que los vecinos más cercanos contribuyan más al promedio que los más distantes (Claudino et al., 2019). De esta forma memoriza la hoja de datos que posteriormente se utiliza como conocimiento para la fase de predicción.

En el ejemplo de la figura 3, para clasificar un objeto nuevo (signo de pregunta), con $k=3$ vecinos, los vecinos más cercanos serían 2 de la clase R (roja) y uno de la clase A (azul). Por mayoría, el objeto se clasifica como clase R.

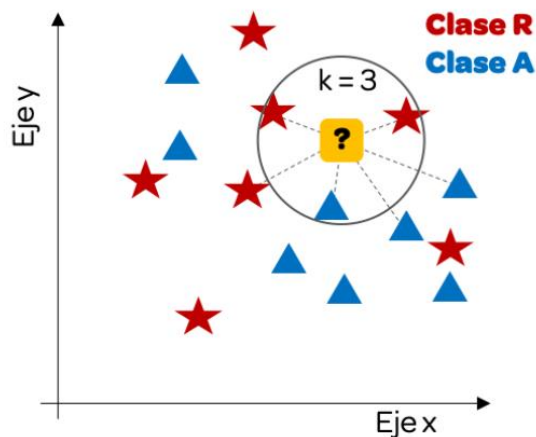


Figura 3. Ejemplo de algoritmo kNN para un $k=3$ (Paredes-Inilupu, 2020).

La técnica ha sido utilizada, por ejemplo, para predecir el resultado de los juegos de acuerdo a la localía del partido a través de indicadores de rendimiento (basados en el comportamiento colectivo de los pases y en función del historial de actuaciones) de los equipos en la Bundesliga de Alemania, encontrando una precisión del 80% de las victorias del equipo local, del 60% de las victorias del equipo visitante y un 20% de los empates (Cintia et al., 2015).

3.2.1.1.2. Árboles de decisión

Los árboles de decisión o “decision tree” (DT), pertenecen al enfoque de programación “divide y vencerás”, el cual es utilizado en inteligencia artificial para la toma de decisiones (Rokach & Maimon, 2007). El DT permite la representación gráfica de una serie de reglas sobre las decisiones tomadas para asignar un valor de salida a una determinada entrada. Su estructura es similar a un diagrama de flujo (Figura 4), donde un nodo representa una característica o atributo, la rama representa una regla de decisión y cada nodo u hoja representa el resultado, de forma tal que se divide el conjunto de datos en subconjuntos más pequeños. Así, se va construyendo el árbol, repitiendo el proceso recursivamente para cada atributo, hasta que todas las variables pertenezcan al mismo valor del atributo (Gervilla García & Palmer Pol, 2009).

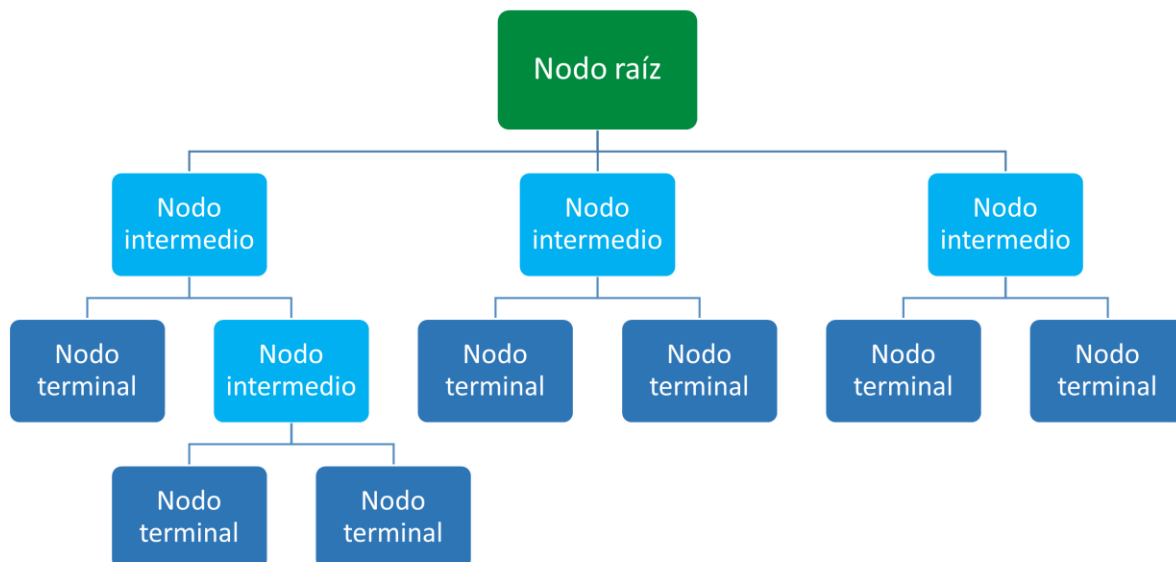


Figura 4. Ejemplo de estructura básica de un árbol de decisión

En la figura 4, se pueden distinguir 3 tipos de nodos: a) el nodo raíz, que significa el conjunto de datos que se dividirán, b) el nodo intermedio, que se encarga de dividir los datos en dos grupos, y c) el nodo terminal, que no se divide más y es donde recae la decisión final. La profundidad tiene que ver con el número máximo de nodos de una rama. Esta condición le permite al modelo ser de gran comprensión para usuarios ajenos a la minería de datos, ya que es fácil visualizar e interpretar. Entre sus desventajas se puede destacar que los datos sensibles al ruido y una pequeña variación en los datos puede dar lugar a un árbol de decisión diferente.

Según el estudio de Claudino et al. (2019), este algoritmo ha sido muy utilizado para predecir el rendimiento y el riesgo de lesiones en los deportes colectivos como fútbol, fútbol americano, volibol, baloncesto, y beisbol, entre otros. Por ejemplo, Rossi et al. (2018), proponen un enfoque multidimensional totalmente basado en datos de GPS recopilados automáticamente para medir la carga de entrenamiento y ofrecer un pronosticador de lesiones preciso e interpretable, brindando un manual simple de reglas para evaluar e interpretar la compleja relación entre el riesgo de lesiones y la carga de entrenamiento.

3.2.1.1.3. Bosques aleatorios

Medina Merino y Ñique Chacón (2017) comparan el siguiente algoritmo como una extensión de los árboles de clasificación. El algoritmo de bosques aleatorios o “Random

Forest” surge como la agrupación de varios árboles de clasificación; cuantos más árboles tenga, más robusto es el bosque. Básicamente selecciona de manera aleatoria una cantidad de variables con las cuales se construye cada uno de los árboles individuales. Las predicciones serán ponderadas a través del cálculo de la clase más votada de los árboles que se generaron, para finalmente hacer la predicción por bosques aleatorios (Figura 5).

Este proceso reduce la tendencia al sobreajuste y la varianza en comparación con los árboles de decisión regulares y, por lo tanto, es una poderosa herramienta para la predicción (Groll et al., 2018). Estos investigadores compararon tres enfoques de modelado diferentes: a) bosques aleatorios, b) modelos de regresión de Poisson, y c) métodos de clasificación. Su propósito era clasificar los resultados de partidos en fútbol con respecto a sus desempeños predictivos basados en todos los partidos disputados durante las Copas Mundiales de FIFA 2002, 2006, 2010 y 2014, encontrando que el método de mayor desempeño fue bosques aleatorios y que al incorporar parámetros de habilidad del equipo como una covariable adicional el poder predictivo aumentó significativamente.

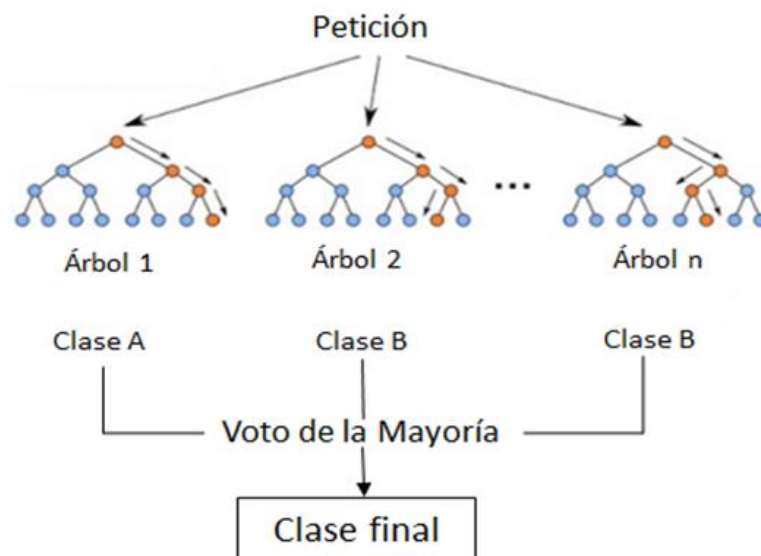


Figura 5. Ejemplo de estructura de bosques aleatorios.

El modelo también fue aplicado durante la Copa Mundial de FIFA 2018 y según las simulaciones, España y Alemania (Campeón Mundial en la edición) resultaron ser los máximos favoritos para ganar el título.

Eryarsoy y Delen (2019) concluyeron en su estudio que el mejor rendimiento para la predicción, según el análisis de sensibilidad y precisión lo brindó el Random Forest, entre otros métodos utilizados (e.g. kNN o máquina de soporte de vectores), alcanzado una precisión de 76% en la clasificación según el resultado del partido (ganar/empatar/perder) y de un 86% en la clasificación según los puntos logrados (puntos/no puntos) durante un estudio longitudinal de 10 años (2007-2017) en la Superliga de Turquía. Entre las variables más importantes destacaron el porcentaje de puntos posibles ganados, la clasificación actual en la liga, y la consistencia del equipo en la liga durante todo este período, al igual que la consistencia en la formación de juego durante la temporada.

3.2.1.1.4. La máquina de soporte de vectores

La máquina de soporte de vectores (SVM), se utiliza normalmente para el reconocimiento de patrones, la clasificación y análisis (Borges, 1998). Es un clasificador discriminativo definido formalmente por una línea de separación llamada hiperplano. En un espacio bidimensional, este hiperplano es una línea que divide al plano en dos partes en las que cada clase se encuentran a cada lado (Figura 6).

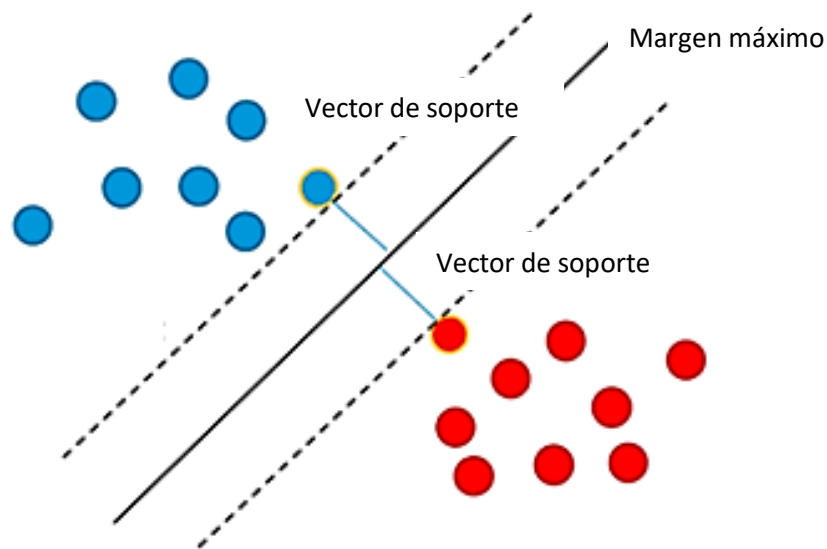


Figura 6. Esquema de un hiperplano (Martinez, 2019).

Los vectores de soporte son los puntos de datos más cercanos al hiperplano. La distancia entre la línea y los vectores de soporte se conoce como margen. A mayor margen, mejor clasificación. Ahora bien, puede ser que los datos se encuentren mezclados entre las

clases y no es posible separarlos con una línea por lo que proyecta los datos a un nuevo eje, a partir de una función llamada kernel (Figura 7).

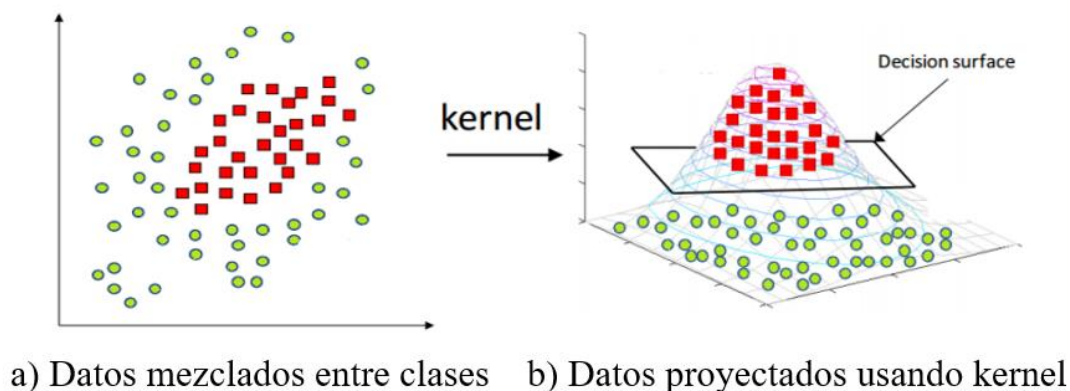


Figura 7. Uso de la función kernel (Martinez, 2019).

Eggels et al. (2016), utilizaron SVM con la función kernel, para explicar los resultados de los partidos de fútbol a partir de la evaluación objetiva de la calidad y cantidad de oportunidades de gol por jugador y logran tener una buena estimación de predicción. El estudio presenta la ventaja de que conduce a información sobre el accionar del jugador para realizar ajustes en su preparación o incluso en la adquisición de nuevos jugadores.

Otro estudio en el que utilizaron SVM y que llama la atención por su alta precisión de predicción (> 90%) fue desarrollado por Gu et al. (2019) en juegos de hockey. Esta técnica confirmó la idoneidad del conjunto de variables elegidos para la predicción del juego y clasificó a partir del desempeño histórico como entrada. Según los autores se logró validar el uso de las 19 métricas “verdaderas” en el entrenamiento, entre las que destacaron los goles, asistencias, % de intentos de tiros (i.e., al arco, fallados o bloqueados), entre otras.

3.2.1.2. Aprendizaje supervisado de regresión

El aprendizaje supervisado de regresión es un subcampo del aprendizaje supervisado cuyo objetivo es establecer un método para la relación entre un cierto número de características y una variable continua. En el deporte, destacan las técnicas de análisis estadístico como la regresión logística, y la regresión múltiple.

3.2.1.2.1. Regresión logística

Es uno de los algoritmos de ML más simples y más utilizados para la clasificación de dos clases. Utiliza una función sigmoide, que se representa en una curva en forma de “S” que puede tomar cualquier número de valor real y asignar a un valor entre 0 a 1 (Figura 8).

Si la salida de la función sigmoide es > 0.5 ; es decir, su punto de inflexión, se puede clasificar el resultado como 1, pero si es < 0.5 se clasifica el resultado como 0. Si el resultado es, por ejemplo, 0.75, se puede afirmar en términos de probabilidad de que hay un 75% de probabilidad de que suceda el evento.

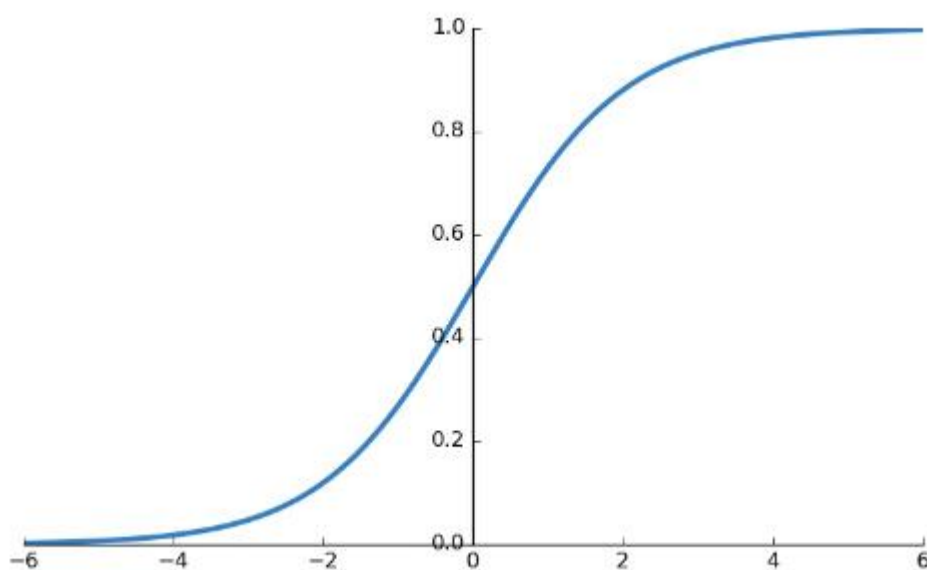


Figura 8. Función sigmoide.

La regresión logística difiere de la lineal al proporcionar una salida discreta y no continua como en el caso de la lineal. Taylor et al. (2008), utilizaron la regresión logística para desarrollar un modelo que evalúa la frecuencia de las actividades técnicas con el balón realizadas por un equipo profesional de fútbol británico en función de la ubicación del partido y la calidad de la oposición con base en el resultado del partido. Enfatizan en la necesidad de que los analistas y entrenadores consideren los posibles efectos independientes e interactivos de la ubicación del partido, la calidad de la oposición y el estado del partido al evaluar los componentes técnicos del rendimiento futbolístico.

3.2.1.2.2. Regresión lineal múltiple

Otros modelos de regresión se han utilizado en el deporte como por ejemplo el modelo de regresión múltiple y es definido por Boada y Mayorca (2011) como un procedimiento poderoso y flexible, utilizado principalmente para analizar las relaciones de asociación entre una variable dependiente métrica y una o más variables independientes o con multicolinealidad reducida. La multicolinealidad es un problema por la gran cantidad de datos derivados y necesita ser controlada para evitar conclusiones erróneas (Springham et al., 2020).

3.2.1.2.3. Regresión por componentes de mínimos cuadrados parciales

La técnica de regresión por componentes de mínimos cuadrados parciales o PLSCA (“por sus siglas en inglés, Partial Least Squares Correlation Analysis”) también ha demostrado solución al problema de multicolinealidad en la regresión múltiple (Weaving et al., 2019). Este modelo ha logrado, por ejemplo, identificar con éxito las variables predictoras para el desarrollo de la “aptitud” en jugadores de rugby profesionales en los índices de carga de entrenamiento.

Springham et al. (2020) también lograron identificar variables predictoras del rendimiento físico en partidos, utilizando el algoritmo de métodos multivariantes con selección de variable imparcial (“MUVR”). Este algoritmo es destinado a encontrar asociaciones entre los datos predictores (una matriz “x”) y una respuesta (un vector “y”) a través del modelado de mínimos cuadrados parciales. MUVR es útil para manejar datos que tienen un gran número de variables y pocas observaciones, y construye modelos multivariados robustos y parsimoniosos que generalizan bien, minimizan el sobreajuste y facilitan la interpretación de los resultados (Shi et al., 2019).

3.2.2. El aprendizaje no supervisado

El aprendizaje no supervisado tiene como objetivo descubrir estructuras y patrones en datos sin etiqueta, basándose únicamente en datos de entrada y sin la intervención humana. Se utiliza para problemas complejos, con una respuesta desconocida. Su función principal es la agrupación de los datos, por lo que el algoritmo debería catalogar por similitud y poder crear grupos (i.e., conglomerados o “clusters”), sin tener la capacidad de definir cómo es cada individualidad de cada uno de los integrantes del grupo.

3.2.2.1. Redes neuronales artificiales

Las redes neuronales artificiales o ANNs (“Artificial Neural Network” por sus siglas en inglés, forma parte de las técnicas de aprendizaje no supervisado más utilizadas. Se agrupan dentro de las técnicas conexionistas de la inteligencia artificial y constituye una de las áreas de estudio más ampliamente difundidas. Las ANNs son capaces de aprender las características relevantes de un conjunto de datos para luego reproducirlas en entornos ruidosos o incompletos, siendo especialmente útiles para tareas de clasificación y regresión (Hammer, 2003). Usualmente, las ANNs reciben la información proveniente del exterior mediante un conjunto de neuronas de entrada y cuentan con un conjunto distinto de neuronas de salida para manejar los resultados.

El uso de redes neuronales artificiales ha aumentado considerablemente en años recientes para la identificación, clasificación y predicción del rendimiento en el fútbol (Hassan et al., 2020). En una revisión sistemática de literatura realizada por Claudino et al. (2019), cerca del 36% de los estudios tomados en cuenta, usaron esta técnica, de los cuales un 10% fueron investigaciones de riesgo de lesión y un 26% en análisis de rendimiento deportivo.

Algunos ejemplos de aplicación de las ANNs en el rendimiento deportivo han sido por ejemplo, la capacidad de discriminar con éxito la eficacia de atributos físicos y técnicos durante el partido de fútbol, cuando se gana o se pierde, con un porcentaje de sensibilidad de 83.3% para cuando se gana y 72.7% para cuando se pierde (Hassan et al., 2020). En otro estudio similar, Huang y Chang (2010), construyeron un modelo para predecir la tasa de victoria de dos equipos durante el mundial de fútbol de Alemania en el 2006 a partir de datos estadísticos oficiales de la etapa de clasificación al evento, con una precisión de 76.9%. Esta técnica también se ha utilizado para analizar la posición de los jugadores (Memmert et al., 2017).

3.2.2.2. Análisis de componentes principales

El análisis de componentes principales (PCA, por sus siglas en inglés), es otro método estadístico que puede proveer una solución en el área de las ciencias del deporte. Este método selecciona las variables significativas y descarta los componentes menos importantes del análisis (Parmar et al., 2018). El tener un gran número de variables predictoras no implica necesariamente mucha información ya que esta puede ser redundante. El PCA precisamente

elimina esta información redundante, por lo que optimiza el proceso de análisis de las variables representativas de desempeño que pueden ser obtenidas (Muazu Musa et al., 2016).

Lago-Peñas et al. (2017), mediante un análisis de regresión y análisis factorial, identificaron y midieron diferentes estilos de juego en el fútbol profesional de China (Superliga China de Fútbol) durante la temporada 2016. Los investigadores pudieron analizar 240 partidos y un total de 20 variables (14 de ataque y 6 de defensa). El uso de diferentes indicadores de desempeño permitiría obtener relaciones de mayor orden entre ellos, y con ello una mejor descripción e identificación de los patrones de juego que pueden reflejar un estilo de juego específico. De hecho, de las 20 variables se obtuvieron cinco factores que tuvieron autovalores mayores a 1 y explicaron el 79.6% de la varianza total:

- a. Factor 1: estilo de posesión de juego, correlacionado con la posesión del balón, la posesión del balón en campo rival, ataques posicionales, pases precisos, pases adelante y atrás.
- b. Factor 2: ataque a balón parado, correlacionado positivamente con el número de ataques a balón parado y ataques.
- c. Factor 3: juego de contraataque, correlacionado con intercepciones, intercepciones en campo contrario, balones recuperados y número de contraataques.
- d. Factor 4 y 5: juego de transición, correlacionado con pérdidas de balón, y balones perdidos en campo contrario.
- e. Factor 5: juego de transición: recuperar balones libres y recuperar balones libres en zona adversaria.

Los investigadores concluyeron que estas métricas pueden permitir a los entrenadores y analistas clasificar los equipos en perfiles específicos de estilos de juego.

En síntesis, se han utilizado diferentes técnicas en el modelado predictivo, encontrando variedad en los resultados, lo cual sugiere que tanto las variables escogidas, como el torneo disputado, el momento de la temporada, o el estilo de juego son condicionantes de la técnica seleccionada. Aún hoy en día, el conocimiento se encuentra en fase exploratoria, por lo que no hay claridad sobre cuál será el mejor método de análisis (Eryarsoy & Delen, 2019; Moncada-Jiménez et al., 2021).

4. Análisis de la actividad del jugador de fútbol según su posición

Es indiscutible que la naturaleza del juego de fútbol es verdaderamente compleja de analizar, ya que se deben considerar las dimensiones técnicas, tácticas y físicas que demandan durante el mismo. La dinámica del movimiento dentro del contexto del partido produce miles de valores diferentes e irrepetibles (Bradley & Ade, 2018).

Los estudios han dado más o menos importancia a la consecución del resultado a partir de las variables físicas o a partir de las variables técnico-tácticas. De acuerdo con algunos estudios, las habilidades técnicas-tácticas son las mejores predictoras del éxito (i.e., ganar) (Asian Clemente et al., 2019; Carling et al., 2007; Castellano Paulis et al., 2012; Longo et al., 2019) y son discriminadores notables entre estándares competitivos (Lago-Peñas & Dellal, 2010; Liu et al., 2016; Sarmiento et al., 2014). Otros enfatizan en la importancia del rendimiento físico como determinante del resultado final del partido (Andrzejewski et al., 2018; Dellal et al., 2011; Konefał, Chmura, Kowalczyk, et al., 2019; Konefał et al., 2020; Mackenzie & Cushion, 2013).

Bradley y Ade (2018), desde una perspectiva más integral, invitan a analizar el fútbol como un deporte multifacético con aspectos físicos, técnicos y tácticos que se fusionan para influir en el rendimiento, sin excluir un factor de otro. De hecho, afirman que la táctica o el contexto son los principales moduladores del rendimiento físico.

Precisamente al analizar cualquier conjunto de datos será necesario tener en cuenta los factores situacionales que condicionan la estrategia del equipo y jugador durante el partido (Lord et al., 2020; Sarmiento et al., 2014; Taylor et al., 2008). Algunos de estos factores son por ejemplo, la ubicación del juego (Aquino et al., 2020; Oliva-Lozano et al., 2021; Rampinini, Coutts, et al., 2007; Taylor et al., 2008), la calidad de la oposición (Lago-Peñas et al., 2010; Lago-Peñas & Dellal, 2010; Rago et al., 2021; Rampinini, Coutts, et al., 2007), el momento del juego (e.g., ir ganando o ir perdiendo) (Konefał, Chmura, Kowalczyk, et al., 2019).

Otra condición que varía la respuesta de los jugadores durante el partido y que para efectos de la presente propuesta será importante desarrollar, es indudablemente la posición y el rol que debe cumplir un jugador dentro del campo, ya que depende de la formación de juego que se utilice (e.g., 1-4-4-2 o 1-4-2-3-1) y de la posición específica en el terreno de juego requerirá diferentes esfuerzos físicos y diferentes exigencias técnico-tácticas (Arjol-

Serrano et al., 2021; Borghi et al., 2021; Dellal et al., 2011; Konefał, Chmura, Kowalczyk, et al., 2019).

Por ejemplo, Arjol-Serrano et al. (2021), compararon dos formaciones de juego, la 1-4-2-3-1 y la 1-4-4-2, e indican que las posiciones específicas de los jugadores analizadas, requieren diferentes exigencias físicas en los centrodelfanteros y mediocampistas centrales, mientras que en las acciones técnico-tácticas los cambios fueron más marcados en los laterales, mediocampistas centrales, mediocampistas abiertos y los volantes ofensivos. También han encontrado diferencias significativas entre el sistema 1-3-5-2 comparado con sistemas de cuatro defensores (e.g., 1-4-3-3 o 1-4-2-3-1), especialmente en variables físicas como distancia total, alta intensidad y alta carga metabólica (Tierney et al., 2016).

De acuerdo con algunos investigadores, para mejorar el éxito del equipo se requiere un mayor nivel de actividad física de los jugadores en algunas posiciones de la cancha y una mayor actividad técnica de los jugadores en otras (Di Salvo et al., 2013).

A continuación, se analizarán los indicadores que han tenido mayor impacto para determinar el rendimiento de los jugadores durante los partidos. Es necesario aclarar que se estudiarán las características individuales, tomando en cuenta las variables tácticas, técnicas o físicas que describen todos los eventos de situaciones de “1 vs. 1” durante el juego ofensivo y defensivo con y sin balón; sin tomar en cuenta la táctica grupal, que representa la cooperación entre subgrupos (e.g., el bloqueo defensivo durante una trampa de fuera de juego) o la táctica de equipo que refiere a la formación del equipo ofensiva y defensivamente.

4.1. Variables técnico-tácticas

Son muchas variables técnicas que han sido utilizadas en los estudios; sin embargo, se coincide que entre las más relevantes se consideran los tiros, centros, pases, posesión del balón y jugadas de 1 vs. 1 (Castellano Paulis et al., 2012; Konefał, Chmura, et al., 2019a; Lago-Peñas et al., 2011; Liu et al., 2016; Szwarc et al., 2017), ya que su alta efectividad es correlacionada con el resultado del partido (Konefał, Chmura, Kowalczyk, et al., 2019; Link & de Lorenzo, 2016). Para Lago-Peñas et al. (2011), estas son funciones discriminantes que clasificaron correctamente el 79.7% de los triunfos, empates y pérdidas durante 288 juegos de la UEFA Champions League en tres ediciones (2007-2008, 2008-2009, 2009-2010).

4.1.1. Variables de ataque

4.1.1.1. Gol/asistencia

Durante la Copa Africana de Naciones (AFCON) 2017, un estudio examinó los indicadores de rendimiento de los partidos que discriminaban entre equipos ganadores, los que empataban y los perdedores. Se recopilaron datos de 32 partidos y se incluyeron variables como el número de goles marcados, el período de tiempo en el que se marcó un gol y el impacto del primer gol en el resultado del partido, entre otras. Los resultados mostraron que los goles marcados (1.80 ± 0.83) fueron el indicador discriminativo del rendimiento de los equipos ganadores (Kubayi & Toriola, 2020). Este hallazgo es algo intuitivo, ya que a mayor cantidad de goles marcados existe una mayor probabilidad de ganar un juego.

Eggels et al. (2016), proponen un método para determinar el ganador esperado de un partido de fútbol de élite, estimando la probabilidad de anotar las oportunidades de gol individuales. De acuerdo con los investigadores, los goles son extraños en el fútbol, en el sentido de la gran influencia de un solo gol en el resultado de un partido y consideran que se obtiene una forma más objetiva de analizar el resultado a través de las oportunidades de gol en lugar de los goles reales marcados. Para su análisis toman en cuenta variables como la distancia a la portería, zona del cuerpo con la que se ejecuta el remate, ángulo de remate, origen de la acción, calidad del jugador ofensivo, y la calidad del portero, entre otras.

4.1.1.2. Pase

En la literatura se han descrito diferentes indicadores de pases, dependiendo de la cantidad, la precisión, la longitud (e.g., cortos, medios, largos), la dirección (e.g., adelante, atrás, al lado), el origen (e.g., mitad propia o mitad contraria), y han estado vinculados con el éxito de los equipos o clasificando el nivel competitivo de los jugadores (Barron et al., 2020; Konefał, Chmura, Kowalczyk, et al., 2019). La cantidad de pases ha sido la variable que más ha tenido seguimiento por parte de los científicos y es una de las que ha sido muy palpable la evolución que ha tenido en el fútbol en los últimos 45 años, ya que se han visto aumentos en la tasa de pases mayores de 35% cuando se compararon los juegos finales de la Copa del Mundo de 1966 al 2010 (Kempe et al., 2016), o incluso de un 40% en la Liga Premier de Inglaterra al comparar siete temporadas consecutivas desde el 2006 hasta el 2013 (Bush et al., 2015). Este aumento también se ha asociado a la capacidad de los equipos para ganar los partidos (Barron et al., 2020; Konefał, Chmura, Kowalczyk, et al., 2019).

Barron et al. (2020), con jugadores de la Liga Premier inglesa, lograron construir 14 modelos predictivos, todos con una precisión $\geq 75\%$, a partir del uso de ANNs, categorizando tres estatus diferentes de jugadores a partir de la posición de juego. Los investigadores encontraron dentro de las variables predictoras, las relacionadas con la capacidad de pase (48 de las 134 variables medidas) fueron las más importantes. Destacan además los pases acertados, los pases hacia adelante y el origen del pase en campo rival.

En otro estudio, Hassan et al. (2020), encontraron que el pase de mediana distancia fue la variable que más peso tuvo en su modelo predictivo, que alcanzó un 83.3% de sensibilidad para la victoria durante la Copa Mundial de fútbol de Rusia 2018.

Cintia et al. (2015), elaboraron un indicador de rendimiento basado en el comportamiento colectivo de los pases (“H”), que consistió en la suma de varios parámetros asociados, entre ellos volumen total de pases del equipo, volumen medio de pases entre jugadores, variación del volumen de pases de los jugadores, volumen medio de pase en zonas del campo, y variación del volumen de pases en las zonas del campo. Informaron que en el 73% de las victorias en casa, el equipo local tuvo un indicador H más alto que el equipo visitante, mientras que en el 51% de las derrotas en casa, el equipo tuvo un indicador H más bajo que el oponente de visita.

Es importante también tomar en cuenta que el pase y sus variaciones dependerán del estilo de juego que cada entrenador implemente, ya que según Sarmiento et al. (2014), secuencias de pase más largas pueden incluso disminuir la probabilidad de remates al marco del equipo contrario. Dentro de esta categoría se incluye el centro, y se define como un lanzamiento del balón al área de penal del adversario desde el costado y ha sido identificado como un factor clave para diferenciar equipos exitosos y no exitosos (Geurkink et al., 2021; Lago-Peñas & Dellal, 2010; Lago-Peñas et al., 2011). Sin embargo, también se ha etiquetado como un método ineficaz para crear oportunidades de gol (Liu et al., 2016).

4.1.1.3. Remate o Tiro al marco

Son todos los tiros orientados hacia la portería rival con el objetivo de anotar un gol. Las investigaciones comúnmente han utilizado como predictores de éxito, el tiro total (i.e., cantidad de remates a portería) o la precisión en el remate (Kubayi & Toriola, 2020; Lago-Peñas & Dellal, 2010). También se han estudiado los tiros por fuera de la portería, los tiros al poste-larguero y los tiros bloqueados (Castellano Paulis et al., 2012; Liu et al., 2016).

Geurkink et al. (2021), construyeron un modelo de aprendizaje automático predictivo, con una precisión de 89.6%. Utilizaron una amplia gama de indicadores de rendimiento ($n = 100$), recolectados durante 576 juegos de la primera división de Bélgica y encontraron que los juegos que resultaron en victoria o en pérdida pudieron clasificarse correctamente, siendo los tiros a portería dentro del área de penal el mejor predictor de todas las variables.

4.1.2. Variables defensivas

Dentro de las variables defensivas que han estado asociadas al éxito se pueden encontrar las intercepciones (i.e., acción de recuperación del balón que interrumpe el ataque rival), los despejes (i.e., golpe del balón por parte un jugador defensor, orientado hacia zonas alejadas de la propia portería) y los desafíos aéreos (i.e., disputa de dos rivales por el balón que se encuentra más arriba del nivel del hombro) (Taylor et al., 2008).

Los duelos también han sido estudiados y son aquellas acciones de 1 vs. 1 que se presentan cuando el balón es dividido; es decir, que no pertenece a alguno de los dos equipos. Link y de Lorenzo (2016) afirman que la actividad técnica clave que realizan los futbolistas alemanes consiste en duelos, revelando un aumento de un 16% en el número de duelos durante la fase final de la temporada de la Bundesliga. Sin embargo, Konefał, Chmura, et al. (2019a), encuentran un resultado muy interesante de una relación inversamente proporcional entre el número de duelos y el porcentaje de duelos ganados. Los investigadores identificaron una disminución constante en la cantidad de duelos (durante tres temporadas continuas), independientemente de las posiciones de los jugadores en el campo o el resultado del partido. Esto puede deberse a una mejor toma de decisiones por parte de los jugadores y una mayor conciencia de la elección de la actividad técnica y colectiva que tiene más posibilidades de ser efectiva que llevar el juego a lo individual. Al mismo tiempo, el porcentaje de duelos ganados aumentaba en consonancia con posiciones cada vez más ofensivas.

También ha sido clave el factor recuperación del balón en campo rival, que demuestra una reacción defensiva después de perder la posesión del balón (Vogelbein et al., 2014) o recuperar la posesión en el último tercio de la cancha (Almeida et al., 2014), patrones defensivos muy característicos del estilo de juego moderno.

4.2. Rendimiento físico

La distancia total recorrida, la distancia recorrida a alta intensidad, el número de sprints realizados, el promedio de la velocidad en la carrera, la velocidad máxima de carrera

y la carga, son algunas de las variables más estudiadas (Andrzejewski et al., 2018; Chmura et al., 2018), las cuales se describen a continuación.

4.2.1. Variables relacionadas con distancia

4.2.1.1. Distancia total recorrida

Actualmente, los científicos reportan recorridos entre 9 a 13 km/partido (Bradley & Ade, 2018; Chmura et al., 2017; Rampinini, Coutts, et al., 2007; Vigne et al., 2010). Durante el Campeonato Mundial de Fútbol Brasil 2014, los equipos recorrieron mayor distancia entre los cuartos de final y las semifinales; no así en la final. Incluso se logró comprobar que los jugadores de la Selección Alemana (campeona mundial) recorrieron en promedio, mayor distancia total en relación con la media de otros equipos (10.39 ± 1.19 km vs. 10.05 ± 0.95 km) (Chmura et al., 2017).

Hassan et al. (2020), destaca dentro de los 75 atributos de juego medidos y su relación con el resultado del partido, como la distancia cubierta sin la posesión del balón y el promedio de distancia cubierta por el equipo son poderosos predictores de éxito, durante el Campeonato Mundial de Fútbol Rusia 2018. No obstante, algunos aún consideran que la distancia recorrida por un jugador es una forma bastante superficial de evaluación del juego y que será necesario tener en cuenta la cantidad de carreras de velocidad y otras actividades de intensidad máxima y submáxima (Andrzejewski et al., 2018).

4.2.1.2. Distancia recorrida a diferentes intensidades

Los investigadores han utilizado descriptores genéricos para categorías de movimiento (e.g., trotar, correr, velocidad máxima). También se han creado una serie de umbrales de velocidad (e.g., 0-6 km/h, 6-12 km/h, 12-19 km/h, 19-24 km/h, > 24km/h) debido a las variaciones en el sexo del jugador, la maduración, los estándares competitivos y la capacidad física (Bradley & Ade, 2018; Cummins et al., 2013).

El análisis de las distancias recorridas a diferentes velocidades durante el juego ha acaparado la atención en la literatura ya que se considera como un reflejo de las demandas físicas de la competencia (Chmura et al., 2018; García-Ramos et al., 2018; Geurkink et al., 2021; Konefał et al., 2020). Vigne et al. (2010), en un estudio retrospectivo de medición durante tres temporadas a un equipo exitoso de la Serie A italiana, demostraron que los jugadores a lo largo del período estudiado cada vez recorrieron menos distancias

submáximas, y sugieren que esto se debe a un aumento de las carreras a alta velocidad en la última temporada medida, elevando aún más la demanda física de los jugadores.

Algunos umbrales que se han asociado al éxito han sido, por ejemplo, distancia recorrida de 15-20km/h en los partidos del Campeonato Mundial de Rusia 2018 (Hassan et al., 2020) y en la liga profesional belga (Jupiler Pro League), en donde se observaron distancias recorridas entre 6-15km/h. Estas velocidades se justifican como una necesidad de mantener esfuerzos de intensidad baja a media a lo largo del juego (Geurkink et al., 2021).

Por su parte, Konefał et al. (2020), exponen algunas conclusiones interesantes en cuanto a velocidades submáximas y su relación con el momento del partido:

- a. En intensidad submáxima, por cada minuto en que el equipo va ganando, la distancia recorrida se redujo en 1.1 m comparado con cada minuto cuando el equipo estaba perdiendo.
- b. Cada minuto ganando aumentó en 2.1 m ($p < 0.01$) la distancia recorrida a baja velocidad (11.1 – 14.0km/h) en comparación con cada minuto perdiendo.
- c. Para cada minuto ganando, la distancia recorrida caminando y trotando (0-11 km/h) aumentó en 2.2 m ($p < 0.05$) comparado con cada minuto perdiendo.

Por ende, en la interpretación de las variables físicas se debe considerar que “más” no siempre es “mejor”, como lo demuestran las distancias de carrera en las distintas velocidades submáximas (Geurkink et al., 2021). Posiblemente cuando se está ganando, la estrategia del partido será más defensiva, provocando que el equipo esté más compacto y evite dejar espacios, para salvaguardar más eficientemente la propia portería o incluso puede ser que los equipos altamente exitosos pueden realizar actividades de menor intensidad durante los partidos en virtud de ser técnica y/o tácticamente superiores (Di Salvo et al., 2009).

4.2.1.3. Carrera de alta velocidad (HSR, “High speed running”)

En la literatura existen diferentes umbrales para determinar la carrera de alta intensidad, comúnmente denominada como “High speed running” (HSR). Algunos estudios la conceptualizan como la distancia recorrida a una velocidad entre 21 km/h y 24 km/h (Oliva-Lozano et al., 2021) o carreras que van de 19.9 a 25.2 km/h (Chmura et al., 2017; Malone et al., 2018; Rampinini, Bishop, et al., 2007).

Las HSR representan entre el 5% al 15% del total de la distancia total recorrida en un juego y en la última década se ha observado un incremento de hasta un 40% (Bradley & Ade,

2018; Malone et al., 2018). La HSR está muy correlacionada con el estado de entrenamiento del jugador (Mohr et al., 2003) y es considerada una variable predictora del rendimiento grupal (Andrzejewski et al., 2018; Chmura et al., 2017; Malone et al., 2018; Springham et al., 2020).

La HSR debe ser desarrollada durante la pretemporada y estar en exposición de carga constante durante los entrenamientos en fase competitiva (Malone et al., 2018), pues se ha demostrado que un óptimo desarrollo de esta cualidad influirá en la capacidad de generar sprints durante los partidos (Springham et al., 2020). También proporcionará protección para evitar lesiones musculares, especialmente para aquellos jugadores que tienen poca participación en los partidos (Malone et al., 2018).

4.2.1.4. Velocidad relacionada con Sprint

Esta cualidad es definida como el número de sprints ejecutados por un jugador a una velocidad > 24 km/h y la respectiva distancia recorrida a esa velocidad (Konefał, Chmura, et al., 2019b). Es una de las acciones más importantes en el juego del fútbol y sólo representa el 1-12% de la distancia total recorrida durante el partido (100-700 m) Se realizan cada 90-180 s y las carreras de velocidad larga tienen una duración promedio de entre 2-4 s (Andrzejewski et al., 2018). Los jugadores de fútbol de élite realizan en promedio 10 sprints por partido, y se alcanzan velocidades superiores a 30 km/h (Oliva-Lozano et al., 2020).

La velocidad máxima promedio, también ha sido reconocida como un indicador de rendimiento, el cual ha evolucionado con el fútbol. Se conceptualiza como la media de la velocidad alcanzada cuando el atleta se desplaza sobre 24 km/h. Durante la Copa Mundial de Fútbol de 2010 en Sudáfrica, los jugadores españoles (campeones del torneo) alcanzaron un promedio de velocidad máxima de 25.52 km/h, mientras que en la edición de Brasil 2014 los jugadores alemanes alcanzaron los 27.89 km/h; es decir, hubo un aumento de 2.4 km/h, sugiriendo un desarrollo en esta cualidad (Konefał, Chmura, Kowalczyk, et al., 2019). Konefał, Chmura, et al. (2019b), propusieron un modelo que predice que cada aumento del 1% en esta variable teóricamente se asociará con un aumento del 4.08% en las probabilidades de ganar en temporadas posteriores.

4.2.2. Variables relacionadas con la carga

Las variables relacionadas a la carga se suelen convertir en valores compuestos para medir carga aguda. Por ejemplo, -7D de carga promedio sería el análogo a la “fatiga” del

jugador. Por su parte, la carga crónica, sería $-28D$ de carga promedio o el análogo a la “aptitud” del jugador. Esta relación carga aguda/carga crónica, permite describir los patrones recientes en la distribución de la carga (Springham et al., 2020).

Existen una gran cantidad de índices de carga de trabajo disponibles para los profesionales, creando una matriz compleja de toma de decisiones, que es a menudo difícil de interpretar (Weaving et al., 2019). La gran cantidad de datos derivados que se obtienen con estos registros, necesitan ser controlados para evitar conclusiones erróneas.

Un ejemplo de estos índices es el “high metabolic load distance” abreviado en HMLd, que viene siendo la distancia (m) recorrida por un jugador cuando su potencia metabólica está por encima de 25.5 W/kg. Este valor corresponde a una carrera a velocidad constante de 5.5 m/s² o cuando hay actividades de aceleración y desaceleración significativas. Por lo tanto, es una combinación de aceleración, desaceleración, sprint y carrera a alta velocidad (Springham et al., 2020), lo cual proviene del gasto metabólico de la actividad física.

En un estudio llevado a cabo con jóvenes jugadores profesionales ingleses (Springham et al., 2020), se correlacionó la carga previa registrada durante los entrenamientos con el rendimiento físico del partido y se logró concluir que los índices de carga de trabajo se identificaron como variables predictoras, de manera que influían en la capacidad de realizar acciones de alta intensidad como los sprint o las HSR. Se demostró que aplicar cargas agudas altas de HMLd en los días -5,-4,-3 con una puesta a punto en los días -2 y -1 mejoraban considerablemente el rendimiento en HMLd durante los partidos. La notación con el símbolo negativo lo que indica es la cantidad de días previos al juego. Por ejemplo, -5, significaría que son 5 días previos al juego.

4.2.3. Variables relacionadas con aceleración

La mayor parte de los estudios utilizan umbrales establecidos de aceleración (ACC) $>3\text{m/s}^2$ y desaceleración (DEC) $> -3\text{ m/s}^2$ (Andrzejewski et al., 2018; Russell et al., 2016; Varley et al., 2017) ya que tienen un impacto significativo en la carga mecánica y en los indicadores de daño muscular de los jugadores (Oliva-Lozano et al., 2020; Springham et al., 2020). Geurkink et al. (2021), utilizaron umbrales más bajos (ACC: $>2\text{m/s}^2$, DEC: -2m/s^2) con futbolistas profesionales y encontraron un peso significativo de esa variable en el modelo predictivo encontrado.

Se sabe que los jugadores desaceleran a alta intensidad más de lo que aceleran, por lo que se recomienda prestar especial atención a los indicadores de carga mecánica (Oliva-Lozano et al., 2020).

4.3. Análisis por posición

4.3.1. Defensas

4.3.1.1. Defensas centrales

Es necesario resaltar que los defensas centrales (DC) han tenido una evolución en los últimos años en cuanto al aumento en el número de pases que ejecutan durante los partidos (Konefał, Chmura, et al., 2019a, 2019b) y en el fútbol profesional inglés, reportaron incluso un 70% de mejora (Bush et al., 2015).

De acuerdo con la literatura consultada, para ganar partidos, los DC deben estar dispuestos a interactuar con el balón, para generar muchas ocasiones de pase, de tipo corto (Adams et al., 2013) y hacia adelante (Barron et al., 2020; Bradley et al., 2013). El número de duelos y el porcentaje de duelos ganados representa una actividad clave para estos jugadores (Konefał, Chmura, et al., 2019a, 2019b).

Por otra parte, en cuanto a valores físicos, los DC son los que menos distancia total recorren (Di Salvo et al., 2009), menos distancia y menos número en sprints (Andrzejewski et al., 2018; Oliva-Lozano et al., 2020) comparados a los jugadores que participan en otras posiciones. Generalmente las actividades de alta intensidad las realizan cuando deben recuperar o correr contra su portería para minimizar los espacios a sus espaldas (Bradley & Ade, 2018). También deben tener una capacidad esencial para acelerar, por la necesidad de ganar diferentes tipos de duelos con los oponentes (Konefał, Chmura, et al., 2019a, 2019b).

4.3.1.2. Defensas laterales

Los defensas laterales (DL) han demostrado un aumento en el número de centros y número de duelos ganados por partido (Barron et al., 2020; Konefał, Chmura, et al., 2019a, 2019b) que, junto a variables de pase como la cantidad, el porcentaje de acierto y la direccionalidad del pase hacia adelante, se han asociado al éxito en el resultado final del partido.

El perfil de los DL desde el punto de vista físico es de especial atención ya que presentan los mejores puntajes en sprint en comparación con el resto de las posiciones en cuanto al número, distancia, promedio de distancia de sprint, y velocidad máxima por partido

(Oliva-Lozano et al., 2020). Precisamente, la literatura destaca estas variables como predictoras en el resultado del partido (i.e., ganar) cuando estos jugadores presentan altos puntajes en ellas (Andrzejewski et al., 2018; Bradley & Ade, 2018; Bradley et al., 2013; Chmura et al., 2018).

Se han encontrado niveles de ACC y DEC más elevados, tanto en distancia alcanzada como en número de eventos (Oliva-Lozano et al., 2020), justificando perfectamente los resultados de Springham et al. (2020) al destacar el rendimiento de estos jugadores en la variable de carga HMLd, definida como una medida global de rendimiento de alta intensidad.

4.3.2. Mediocampistas

Junto a los defensas centrales, los mediocampistas son los jugadores que han tenido una mayor evolución en el número de pases, encontrando aumentos de hasta un 50% en la Liga Premier durante las temporadas 2006/07 hasta 2012/13. El número de pases y el porcentaje de acierto en el pase, el total de tiros, el porcentaje de acierto en el pase y el número de duelos ganados han sido también elementos que han caracterizado hoy en día a estos jugadores y han estado relacionados con el gane (Bradley et al., 2013; Konefał, Chmura, et al., 2019a, 2019b). Es también vital en estos jugadores la conexión hacia adelante, a través del pase, buscando el juego entre las líneas del equipo rival (Bradley et al., 2013).

Los mediocampistas recorren más distancia total que los defensores y delanteros (Vigne et al., 2010). También recorren distancias a muy alta intensidad (Bradley & Ade, 2018; Konefał, Chmura, Kowalczyk, et al., 2019), lo que provoca una gran carga de trabajo HMLd (Springham et al., 2020).

Es importante también aclarar que los mediocampistas pueden tener perfiles diferentes de acuerdo al sistema táctico que emplee el equipo. Por ejemplo, en el sistema de juego 1-4-3-3 deben recorrer mayor distancia y su carga de trabajo (i.e., HMLd) es mayor (>11% y 14%, respectivamente) que cuando juegan en el sistema 1-4-4-2 (Tierney et al., 2016). También ha habido diferencias en un mismo sistema de juego, ya que por ejemplo en el 1-4-4-2, los mediocampistas centrales realizan menor cantidad de sprints por partido (promedio cinco sprints) en comparación con los mediocampistas que juegan por afuera, o sea, los que se ubican cerca las bandas (Oliva-Lozano et al., 2020; Oliva-Lozano et al., 2023). De hecho, estos jugadores presentan valores muy similares a los defensas laterales, ya que un alto porcentaje en los juegos ganados se asocia a cuando recorren mucha distancia a alta

intensidad y realizan un gran número de sprints (Andrzejewski et al., 2018; Konefał, Chmura, et al., 2019a, 2019b; Vigne et al., 2010).

4.3.3. Delanteros

El número de duelos y porcentaje de duelos ganados se han caracterizado en los jugadores que fungen en esta posición, y también se han relacionado significativamente con la victoria (Konefał, Chmura, et al., 2019a, 2019b), al igual que cuando logran una mayor cantidad y eficiencia en los tiros (Bradley et al., 2013). En cuanto al pase, según Bush et al. (2015), el porcentaje de éxito debe ser $> 70\%$ para ganar.

Llama la atención también como las interceptaciones de balón han sido seleccionadas como indicador clave de éxito en los jugadores que juegan en posiciones ofensivas, en este caso delanteros, y no tanto a los defensas centrales o laterales (Barron et al., 2020). Eso puede deberse a la cercanía que estos jugadores tienen con la portería rival, generando un ataque sorpresivo.

Para los centrodelfanteros, las demandas de carrera de alta intensidad y alta velocidad durante los partidos, son en promedio similares a la de los defensores, pero menores que las de los mediocampistas (Bradley et al., 2013). Difiere a los defensores en la forma en cómo hacen la carrera a alta intensidad ya que frecuentemente la realizan en defensa y corriendo en carrera curva, para canalizar a un oponente con el balón en una dirección, y así permitir que los compañeros de equipo apoyen la presión (Bradley & Ade, 2018). Para Konefał, Chmura, et al. (2019b), una velocidad media de carrera en el segundo tiempo de un partido > 0.1 km/h comparada con la del primer tiempo, se asocia con un 27% de probabilidad de ganar un partido.

4.3.3.1. Extremos delanteros

Los extremos delanteros (ED) se distinguen por la capacidad que tienen en los duelos, buscando el 1 vs.1 por el costado con el objetivo de centrar o buscar recortar hacia el centro. El tiro al marco es uno de los indicadores de rendimiento más importantes para los ED.

Al ganar, se ha encontrado que los ED son los que más distancia a alta intensidad recorren, producto de los movimientos específicos que realizan, ya que cuando defienden deben hacer presión alta o presión tras pérdida y cuando atacan se ocupan de centrar en carrera, hacer regates o atacar espacios (Bradley & Ade, 2018; Bradley et al., 2013;

Springham et al., 2020). Los ED presentan un alto nivel de carga HMLd (Springham et al., 2020).

5. Análisis de rendimiento en el fútbol profesional de Costa Rica

En el ámbito profesional del fútbol costarricense en los últimos años se han implementado algunas herramientas para el análisis de rendimiento de los clubes y de las selecciones nacionales de Costa Rica. A la fecha, los clubes nacionales poseen mucha información, sin embargo, gran parte de ella no ha sido utilizada. Según Moncada-Jiménez et al. (2021) en Costa Rica, el aprovechamiento de las herramientas de análisis ha sido escaso, debido a que los clubes de fútbol no cuentan con personal especializado en el procesamiento, la reducción y el análisis de datos.

El acceso a esta información por parte de los científicos de Costa Rica también es complicado, ya que muchas veces los equipos no comparten la información. Esto es evidente en el número de estudios acerca del tema que se pudieron encontrar en las diferentes bases de datos.

En el año 2020, un grupo de investigadores y estudiantes de posgrado de la Universidad de Costa Rica, integrantes de la Escuela de Educación Física y Deportes, del Centro de Investigación en Ciencias del Movimiento Humano (CIMOHU), de la Escuela de Ingeniería Eléctrica y de la Escuela de Estadística realizaron un proyecto de investigación aprobado por la Vicerrectoría de Investigación de la misma entidad con el nombre “Exploración de modelos de análisis de datos provenientes de tecnología inercial utilizados en el fútbol profesional costarricense” (no publicado). El objetivo del proyecto fue determinar las variables físicas que explican el resultado del partido (i.e., ganar, perder o empatar), considerando los datos de los jugadores presentes en cada partido como un conjunto. Ese estudio fue de naturaleza exploratoria, pues no existían antecedentes en el país para ese tipo de tecnología o de análisis de datos.

En el análisis se utilizaron diferentes algoritmos del aprendizaje de máquinas (e.g., árboles de decisión, SVM, k-NN, Nāive Bayes, y análisis de componentes principales, entre otros) y se aplicaron dos enfoques diferentes: a) desde la estadística contemporánea, y b) desde la perspectiva del análisis de patrones de la ingeniería eléctrica. Con ambos enfoques se identificaron variables relevantes que explican cuando el equipo gana. Por ejemplo, desde la estadística contemporánea se logró un indicador de precisión de 0.84 utilizando la técnica

Näive Bayes y las variables que lograron un alto peso de estimación fueron: el oponente (e.g., equipo contrario), los impactos (e.g., 3-5, 5-8, y 8-100 G), la carga metabólica (e.g., distancia recorrida por un jugador cuando su potencia metabólica es >25.5 W/kg) y los sprints (e.g., distancias recorridas sobre el umbral de velocidad absoluto de sprint = 24 km/h).

Mientras tanto, desde la perspectiva del análisis de patrones de la Ingeniería Eléctrica, las variables que tuvieron mayor impacto fueron: el marcador contrario (el equipo suele ganar bajo un marcador de 2 a 0 o 2 a 1 a favor), los impactos (e.g., 0-3G), la velocidad relativa de entre un 45% y un 65% y el HSR por minuto (e.g., 3.75-4.55 m/min).

Otro gran paso del estudio exploratorio fue que se redujo la información importante, de entre 3.5% y 6% del total de las variables que se tenían originalmente para el estudio, dando una luz desde el punto de vista práctico para el seguimiento de las variables relevantes que eventualmente pudieran ser manipuladas por miembros del cuerpo técnico del equipo (i.e., preparador físico, entrenador).

En otra serie de estudios se han descrito las demandas físicas de los jugadores costarricenses durante la competencia. Por ejemplo, Morera-Barrantes et al. (2021) destacan diferencias entre las posiciones de juego, en variables como distancia total recorrida (volantes recorren mayor distancia que defensas y delanteros), impactos totales (mediocampistas presentan más impactos totales que defensas y delanteros) y distancia recorrida entre 11.1-14 km/h, así como 14.1-19 km/h (mediocampistas recorren más que defensas y delanteros). También analizan el nivel competitivo de juego a través de la categorización de los jugadores de acuerdo al nivel competitivo de los equipos en los que participaban. En ese estudio, se encontró que los que pertenecían a la Selección Absoluta de Costa Rica mostraban mejores puntajes en aceleraciones, desaceleraciones y distancias recorridas entre 0-11km/h en comparación con jugadores de equipos de la primera división costarricense pertenecientes a los de mejor clasificación en la tabla de posiciones (“élite”) y a los de peor clasificación (“sub-élite”).

Rivas Borbón et al. (2017), realizaron una comparación del rendimiento físico de los jugadores pertenecientes a las selecciones de Costa Rica y Alemania, y su relación con la posición final alcanzada en la Copa Mundial de Fútbol FIFA Brasil 2014. Lograron demostrar que en promedio los jugadores de la selección de Alemania (Campeona Mundial) recorrieron más metros en esfuerzos físicos de mediana (e.g., 13.0-17.9 km/h) y alta

intensidad (> 18 km/h) que los de Costa Rica (8^{vo} lugar). Además, los jugadores costarricenses recorrieron significativamente más distancia al caminar y andar (0-6.9 km/h) o carrera suave (7.0-12.9 km/h) que sus similares alemanes. Concluyen que este comportamiento tendió a observarse entre las selecciones que se ubicaron en los primeros lugares y las que ocuparon los últimos lugares.

Capítulo II. Justificación

Como se ha descrito anteriormente, la evolución que ha tenido el fútbol con la aparición de herramientas de evaluación, así como la evolución en los enfoques estadísticos por parte de especialistas, ha permitido a los entrenadores dirigir a sus equipos con técnicas cada vez más especializadas. En los últimos 20 años en Costa Rica, ha existido un avance más influenciado por la implementación de metodologías importadas y poco a poco la introducción de nueva tecnología ha permitido dar un enfoque distinto y más profesional en los procesos de entrenamiento y competencia.

El control de la intensidad, por ejemplo, se basaba casi exclusivamente en la frecuencia cardíaca (FC) y la percepción del esfuerzo. En una primera instancia se le pedía al jugador tomar el pulso a través de la palpación después de un ejercicio determinado. Posteriormente, apareció el uso de cardio frecuencímetros o monitores de FC que registraban la información de forma diferida (e.g., Polar Team®), y que una vez terminada la práctica había que descargar los datos para analizarlos. Luego los datos de FC se lograban obtener en tiempo real y junto con software especiales, se controlaban las diferentes zonas de intensidad por las que el jugador se situaba durante el entrenamiento.

En el año 2018, surgió un instrumento que al menos en Costa Rica, parecía difícil de adquirir por el alto costo económico, como es la tecnología inercial. En aquel momento los indicados a asumir la implementación de este equipo fueron los preparadores físicos, generando en ellos un enorme reto, no sólo por el uso del aparato (i.e., funcionamiento, toma y procesamiento de información) sino también por el análisis de los datos, al encontrar un sinnúmero de variables, que además no eran sencillas de interpretar.

Esto obligó a centrarse en lo que realmente se consideró “necesario”, seleccionando algunas de ellas, y que a la postre, generaron confusión al compararlas con los resultados de los partidos que se disputaban (e.g., distancia total de carrera, ya que muchos partidos se ganaron corriendo poca distancia o se perdieron corriendo mucha distancia), o con datos provenientes de otras latitudes, descontextualizando la realidad local.

También empezó a aflorar el uso de herramientas de video y estadísticas a través de plataformas especializadas, por parte de los entrenadores, que permitieron poder editar acciones de partidos para luego mostrarlas a sus dirigidos con argumentos más objetivos. Sin

embargo, es tanta la cantidad de métricas disponibles en estas plataformas que los mismos entrenadores no toman en cuenta, muchas veces por desconocimiento en las funciones del software especializado o por la saturación de métricas que pueden incluso generar confusión.

Esta situación ha exigido modificar el perfil, que es eminentemente práctico, a una necesidad de enfocarse más a lo investigativo y científico, a la ciencia de los datos, profundizando aún más en las variables situacionales y contextuales del juego, con el fin de tener un criterio más objetivo de lo que realmente sucede durante la competencia.

Este análisis objetivo de los datos permitirá tener un gran soporte para el planeamiento táctico y físico de un equipo que, en un futuro cercano, podrá ser útil para implementarlo, iniciando con divisiones menores como una manera de enfatizar en aquellos aspectos que deben ser abordados de forma más profunda, y entendiendo que a este nivel la implicación que existe es poca ya que en primera división muchas veces no es posible ejecutar. Aunque el fútbol es totalmente impredecible, estas herramientas tecnológicas permiten comprender el comportamiento de las variables que podrían ser posteriormente utilizables.

Al existir una escasez de estudios de este tipo en el medio nacional, y tratando de dar respuesta de cuáles variables son determinantes para aumentar las probabilidades de ganar según la posición de juego, se ha diseñado este estudio.

Capítulo III. Propósito

a. Objetivo General

Analizar retrospectivamente el rendimiento de los jugadores por posición de un equipo de fútbol profesional costarricense a partir de variables físicas y técnico-tácticas recolectadas por medio de tecnología inercial y de la plataforma de datos Instat[®].

b. Objetivos Específicos

1. Determinar las variables que explican el rendimiento de un futbolista en un juego de fútbol según su posición.
2. Identificar las variables que aumentan las probabilidades de ganar un partido según la posición de juego.

Capítulo IV. Metodología

a. Diseño

El diseño de este estudio es retrospectivo (Moncada Jiménez et al., 2023), ya que los datos fueron recolectados durante las Temporadas 2018-2019 y 2019-2020 del campeonato nacional costarricense de fútbol profesional. Es también un estudio inferencial ya que se deduce a partir de una muestra de datos.

b. Participantes

Se utilizaron datos anonimizados de 42 jugadores masculinos profesionales de fútbol. Los participantes pertenecían a un equipo de Primera División del fútbol de Costa Rica. Los datos se recopilaban a lo largo de 4 Torneos del Campeonato Nacional (Invierno 18', Verano 19', Invierno 19' y Verano 2020'), 2 Torneos de la Liga de Campeones de la CONCACAF (edición 2019 y edición 2020), 1 Torneo de Liga CONCACAF Scotiabank 2019.

Los torneos del campeonato nacional, se jugaron de la siguiente manera. Cada torneo se dividió en 22 partidos de la primera fase, donde jugaron todos contra todos (12 equipos en total) en partidos de ida y vuelta (i.e., casa y visita). En la segunda fase, los 4 primeros lugares de la tabla clasificaron a una semifinal en dos llaves (1^o vs. 4^{to} y 2^{do} vs. 3^{ro}) y el ganador de cada llave asistía a una final de la fase. Como premio al equipo más regular de la primera fase, se le otorgaba el derecho a una final de campeonato y si repetía como ganador de la segunda fase automáticamente era declarado campeón del Torneo en disputa.

En cuanto a los torneos internacionales “Liga de Campeones de la CONCACAF” y “Torneo de Liga CONCACAF Scotiabank”, se utilizó un formato de eliminación directa, a través de partidos de ida y vuelta. Los partidos internacionales disputados fueron 12 y mayoritariamente durante el Torneo de Liga CONCACAF Scotiabank (8 en total).

Durante el período competitivo estudiado la formación de juego que se utilizó con mayor frecuencia fue el 1-4-3-3 (Figura 9), con una formación en triángulo en medio campo.

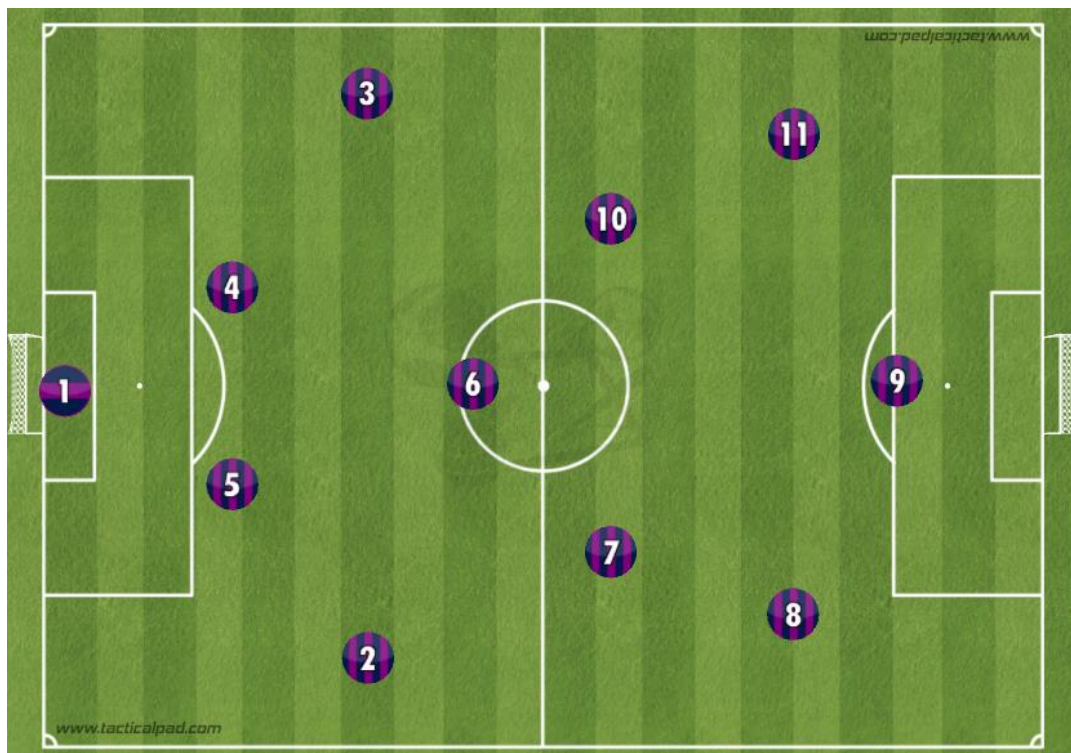


Figura 9. Formación de juego más frecuentemente utilizada de equipo analizado. Jugador #1: portero; jugadores #4,5: defensas centrales; jugadores #2,3: defensas laterales; jugador #6: volante central; jugadores #7,10: Mediocampistas ofensivos; jugadores #8,11; Extremos y jugador #9: centrodelantero

Partiendo de este sistema los jugadores se dividirán de acuerdo a seis posiciones de juego:

- a. 5 defensas centrales (DC)
- b. 9 defensas laterales (DL)
- c. 6 mediocampistas defensivos (MD)
- d. 5 mediocampistas creativos ofensivos (MO)
- e. 9 extremos delanteros (ED)
- f. 8 centrodelanteros (CD)

Como criterios de inclusión, para el análisis se limitó a los jugadores de campo; es decir, se excluyeron los porteros, y además que hubiesen participado durante los partidos completos o al menos durante 70 min de juego.

c. Instrumentos de medición

Para la recolección de datos físicos, se utilizó el dispositivo inercial marca WIMU PRO[®] (RealTrack Systems, Almería, España). El dispositivo está compuesto por un procesador Intel Atom 2 cores de 1 GHz, sistema operativo Linux, GNS, GNSS, un sensor UWB (Ultra Wide Band) con una precisión < 10 cm, acelerómetro de 16G a 1000 Hz, acelerómetro 400G a 1000 Hz, giróscopo a 1000 Hz con registro de hasta 8000 °/s, magnetómetro a 100 Hz, barómetro a 120 kPa, WiFi 802.11 B/G/N, Bluetooth 4.1., ANT+, alta velocidad USB 2.0, NFC, 70 g de peso, 15 mm de tamaño y batería de 1300 mAh.

Esta tecnología permite obtener alrededor de ciento veinte variables, entre las que destacan, por ejemplo, la posición del jugador en el terreno de juego, aceleraciones, impactos, velocidades, tiempo de reacción, saltos, golpes, velocidad angular, rodamientos, inclinación, y dirección de los movimientos (e.g., adelante, atrás, lateral), entre otros. Sin embargo, para efectos del presente trabajo se tomarán en cuenta sólo algunos de ellos (Anexo 1).

La validez y confiabilidad de los datos arrojados por este dispositivo ha sido reportada recientemente. Acciones como carreras de alta intensidad, distancias cortas (10 m) y medias (30 m) o incluso las distancias más largas con variaciones de velocidad, dirección y trayectoria (Bastida Castillo et al., 2018; Muñoz-López et al., 2017), la capacidad de salto de los deportistas (Pino-Ortega et al., 2018), o la distancia total recorrida han sido analizadas, encontrando, por ejemplo, según Bastida Castillo et al. (2018), un sesgo total en la medición de la velocidad entre 1.18 y 1.32 km/h y de la distancia entre 2.32 y 4.32 m para la validez; y buena confiabilidad entre unidades e intraunidades (coeficiente de correlación intraclase, CCI > 0.93), considerando al instrumento como adecuado para mediciones en el fútbol.

Pons et al. (2019) analizan los datos de la demanda de movimiento durante un partido, comparando el Sistema inercial WIMU con Mediacoch[®] (sistema de seguimiento óptico) y concluyen que ambos sistemas presentan una gran concordancia en el coeficiente de correlación intraclase (>0.90 – 1.00) para todas las variables analizadas (i.e., distancia total, distancia por minuto, velocidad media, velocidad máxima, distancia recorrida a diferentes intensidades de velocidad). También el coeficiente de la variación entre dispositivos fue cercano a cero (< 5%) para la distancia total, distancia por minuto, velocidad media, velocidad máxima, andar y trotar, y entre un 9% y un 15% para carrera, carrera intensa y sprint a baja y alta intensidad.

Para la recolección de las variables técnico-tácticas, se utilizó un software de registro con soporte de video, de la marca comercial InStat®. Este software especializado se dedica a evaluar el rendimiento de los equipos y de los jugadores de todas las ligas del mundo, dotando información valiosa desde los aspectos físicos, técnicos y tácticos.

Algunos estudios han demostrado la confiabilidad de sus indicadores de desempeño, a través de la concordancia intraobservador. Se halló un porcentaje de error < 5% (Kubayi & Toriola, 2020) y valores kappa que oscilan entre 0.92 - 0.97 (Casal et al., 2021). Además, recientemente la FIFA lanzó un estándar mundial para sistemas electrónicos de seguimiento de rendimiento (EPTS) (<https://www.fifa.com/technical/football-technology/standards/epts/fifa-quality-performance-reports-for-epts>) e InStat® logró una calificación que lo ubica en el estándar de la industria (<https://digitalhub.fifa.com/m/2fd538ffbae39eb2/original/instat-fifa-epts-report-oct-2019.pdf>).

Logra registrar más de ochenta y cinco variables, que se dividen en estadísticas principales por jugador, entre ellas, minutos jugados, goles, asistencias, jugadas de gol, faltas, tarjetas, tiros a portería, pases, balones perdidos, disputas, rechaces, entre otras. O también, obtener métricas colectivas, por ejemplo, contraataques, posesión de balón, tiros de esquina, tiros libres, etc. En el presente proyecto se tomaron en cuenta algunas estadísticas individuales que se pueden observar en el anexo 2.

d. Procedimientos

Estos datos se recolectaron como parte del trabajo rutinario del preparador físico del equipo (i.e., el tesorero) que es exigido, como parte de su contrato laboral. Todos los jugadores debieron consentir la utilización de los dispositivos como parte del trabajo de campo que se realizaba con el equipo. Verbalmente, los jugadores estuvieron de acuerdo para que los datos pudieran ser utilizados para la investigación, siempre y cuando fueran anónimos.

El dispositivo fue colocado en cada uno de los partidos dentro de un chaleco de neopreno que contiene una pequeña bolsa, ubicada entre las escápulas, donde se inserta el aparato de acuerdo a las especificaciones del fabricante. Este chaleco se utiliza bajo la camisa de competencia oficial, de manera que no es visto por el público y no interfiere con el juego. Cada jugador contaba con su propio dispositivo GPS numerado para evitar confusiones.

Al finalizar el juego, el dispositivo se colocó en una estación que registra los datos y los procesa para luego ser exportados en formato CSV o de hoja de cálculo (e.g., Excel) y ser utilizados para el análisis estadístico.

Toda la recopilación y análisis de datos de carga fue completada por una sola persona (i.e., el tesario). Se recopilaron los datos de todos los partidos competitivos tanto en casa como de visita.

En cuanto a la información obtenida a través de los datos semi-automáticos, se ingresó al sitio web de la plataforma InStat® ([www. https://football.instatscout.com/](https://football.instatscout.com/)) para hacer la búsqueda de los jugadores que participaron durante los partidos de competencia oficiales (nacionales e internacionales) entre las fechas 22/07/2018 al 29/07/2020. Al ser información que proviene de observaciones naturalistas en lugares públicos, no puede considerarse personal o íntima por lo que no requiere del consentimiento informado ni la aprobación de un comité de ética.

e. Análisis estadístico

1. Reducción de datos

Se confeccionó un archivo de datos con 1037 casos de jugadores y 69 variables recopiladas durante los partidos. Para poder llevar a cabo las técnicas de análisis estadísticos propuestas, se necesitaron casos en los que se contara con todas las variables de interés y haber intentado efectuar un análisis con ellas significaba la reducción de la cantidad de casos a la mitad. Esto podría reducir considerablemente la calidad de las estimaciones e inclusive, obtener modelos sobresaturados.

Tomando en cuenta lo anterior, se dividió el archivo de datos según las seis posiciones tácticas, anteriormente mencionadas. Esta diferenciación se realizó con el propósito de distinguir la importancia de las variables según los roles que cumplen los jugadores en el campo y así, conseguir mejores estimaciones.

2. Modelos de análisis

En primera instancia, se confeccionaron modelos de regresión lineal múltiple para asignar una calificación a los futbolistas de acuerdo a los datos registrados durante el partido. De acuerdo con Carrasquilla-Batista et al. (2016), esta técnica se utiliza cuando la variable dependiente “y” depende linealmente de cada una de las variables explicativas x_1, \dots, x_k y un regresor no basta para explicar suficientemente la variabilidad de “y”. En este caso, el

modelo se ajustó utilizando como variable respuesta el Índice InStat[®]. Es importante destacar que este índice de rendimiento se obtiene por medio de un algoritmo que considera la contribución del jugador al éxito del equipo, la importancia de sus acciones, el nivel del oponente y el nivel del campeonato en el que juega. Los cálculos no están disponibles pues sólo los conoce el fabricante de la plataforma (i.e., secreto industrial). La calificación se crea automáticamente y el peso de los factores de acción difiere según la posición; por ejemplo, los errores graves cometidos por los defensores centrales y su frecuencia afectan en mayor medida al índice, que los errores cometidos por los delanteros (Modric et al., 2019).

Bajo el mismo enfoque estadístico, para poder identificar las variables que explican la probabilidad de ganar un juego de fútbol, se ajustaron modelos de regresión logística para cada posición del campo. Barón y Téllez (2004) señalan que si se tiene una variable que describe una respuesta en forma dicotómica (e.g., ganar o perder) y se quiere estudiar el efecto que otras variables independientes tienen sobre ella, el modelo de regresión logística binaria puede ser de gran utilidad para estimar la probabilidad de que se presente el evento de interés, dado los valores de las variables independientes y evaluar la influencia que cada variable independiente tiene sobre la respuesta en forma de una razón de momios o razón de probabilidades, llamada en inglés “odds ratio” (OR). Un valor $OR > 1.0$ indica un aumento en la probabilidad del evento y un valor $OR < 1.0$, implica una disminución. Debido a la poca cantidad de casos completos y la gran cantidad de variables, se decidió aplicar remuestreo de Bootstrap con el fin de mejorar las estimaciones; para lo cual se realizaron 50 remuestreos. Como punto de corte para determinar qué variables eran importantes para predecir el resultado del partido, se consideró aquellas que aparecieron en más del 70% de los modelos ajustados. Con base en lo anterior, se ajustó un modelo final para cada posición.

Además, bajo otra perspectiva, cuando se cuenta con casos en los que se quiere predecir una variable categórica, en este caso dicotómica, también se puede emplear el uso de modelos de ML para clasificaciones. En este caso, se trabajó con las técnicas de árboles de decisiones y bosques aleatorios. Para Bouza y Santiago (2012), la principal ventaja de los árboles de decisiones es que se presentan como una herramienta de clasificación muy potente; además de su facilidad de comprensión para usuarios ajenos a la minería de datos. Esto se debe a que los árboles de decisión permiten la representación gráfica de una serie de reglas

sobre las decisiones tomadas para asignar un valor de salida a una determinada entrada (Gervilla García & Palmer Pol, 2009).

Con respecto a los bosques aleatorios, Hastie et al. (2003), citados por Medina Merino y Ñique Chacón (2017), mencionan que son una combinación entre los árboles de decisión y la técnica del Bagging. Este método trabaja con una colección de árboles incorrelacionados y los promedia, donde cada árbol depende de los valores de un vector aleatorio de la muestra de manera independiente y con la misma distribución de todos los árboles en el bosque.

Para el análisis de clasificación se decidió realizar una dicotomización de la variable respuesta “Resultado”; en donde se pasó de contar con los niveles “Ganar”, “Perder” y “Empatar” a “Ganar” y “No ganar”.

3. Software de análisis estadístico

Para analizar los datos se utilizó el programa estadístico R, versión 4.0.0 (R Core Team, 2020), haciendo uso del entorno de programación R Studio, en la versión 1.1.447. La estadística descriptiva incluye la media y la desviación estándar ($M \pm DE$).

Para los análisis de regresión se utilizaron los paquetes dplyr (Wickham et al., 2018), bootStepAIC (Rizopoulos, 2015), (Wickham et al., 2021), car (Fox et al., 2022), lmtest (Hothorn et al., 2022), patchwork (Pedersen, 2022), tidymodels (Kuhn & Wickham, 2022), rattle (Williams et al., 2022), bitops (Dutky & Maechler, 2022), tibble (Müller et al., 2022), DT (Xie et al., 2023), plotly (Sievert et al., 2022) y caret (Kuhn et al., 2022). La técnica de árboles de decisión se realizó por medio de la biblioteca rpart (Therneau et al., 2022) y la de bosques aleatorios fue desarrollada mediante la biblioteca ranger (Wright et al., 2022). Todos los análisis estadísticos fueron propuestos y realizados por estudiantes avanzados y profesionales en Estadística de la Escuela de Estadística de la Universidad de Costa Rica.

Capítulo V. Resultados

a. Estadística descriptiva

La estadística descriptiva ($M \pm DE$) de los jugadores que participaron en el estudio por posición se presenta en la tabla 1.

Tabla 1. Estadística descriptiva de los jugadores por posición en el campo de juego ($n = 42$).

Posición	Edad (años)	Peso (kg)	Estatura (cm)
Defensas centrales	28.31±4.01	78.52±5.97	182.60±6.10
Defensas laterales	24.15±5.58	74.46±6.04	179.60±6.14
Mediocampistas defensivos	28.58±5.33	72.83±6.54	175.16±2.93
Mediocampistas creativos ofensivos	26.24±4.70	73.47±4.44	175.60±6.22
Delanteros extremos	25.76±5.89	72.58±6.87	174.44±3.88
Centro Delanteros	27.38±6.34	77.52±6.34	180.00±7.19
Todo el equipo	26.49±5.42	74.78±6.23	177.74±5.86

b. Modelos de regresión por posición en el campo de juego

Para los **defensas centrales** se ajustó un modelo basado en 195 casos. Dicho modelo tiene una capacidad predictiva del 76.6%. Se determinó que las variables más importantes que mejoran la calificación del jugador son los goles anotados, las asistencias, los tiros, el porcentaje de disputas defensivas ganadas, las entradas efectivas, los pases efectivos y los pases de finalización efectivos, el porcentaje de acciones efectivas, y las asistencias esperadas. Por otra parte, las que empeoran la calificación son las jugadas de gol, los rechaces, el porcentaje de efectividad en lanzamientos, los tiros por afuera y los tiros interceptados (Tabla 2).

Tabla 2. Variables del modelo de regresión lineal ajustado para defensas centrales.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	113.256	17.800	6.363	<0.0001
Goles	40.957	9.027	4.537	<0.0001
Asistencias	44.350	12.860	3.449	<0.0001
Jugadas de gol	-15.640	5.835	-2.680	0.010
Tiros	46.859	10.857	4.136	<0.0001
% disputas defensivas ganadas	45.469	9.464	4.804	<0.0001
% disputas por arriba ganadas	8.710	5.256	1.657	0.100
Robos de balón con éxito (%)	7.006	5.127	1.366	0.174
Rechaces	-1.008	0.490	-2.056	0.041
Disputas en ataque ganadas	0.055	1.251	1.642	0.102
Entradas efectivas	2.915	1.191	2.448	0.015
Pases efectivos	0.305	0.100	3.050	0.003
Pases de finalización efectivos	14.721	5.374	2.739	0.007
% acciones efectivas	58.373	23.727	2.460	0.015
% efectividad en lanzamientos	-43.781	11.743	-3.728	<0.0001
Tiros fuera	-41.474	11.178	-3.710	<0.0001
Tiros interceptados	-37.388	12.841	-2.912	0.004
Asistencias esperadas	49.071	21.627	2.269	0.024
Torneo Invierno 2019	-21.634	19.742	-1.096	0.275
Torneo Verano 2019	6.603	2.853	2.315	0.022
Torneo Verano 2020	-6.867	19.480	-0.353	0.724
Resultado empate	-6.322	4.352	-1.453	0.148
Resultado victoria	15.656	3.799	4.121	<0.0001

Con respecto a los **defensas laterales**, se tomaron en cuenta 164 casos y se obtuvo un modelo con una capacidad predictiva del 56.24%, la más baja de todos los modelos. Sin embargo, se obtuvieron variables consideradas como significativas que tienen mucho sentido ya que cumplen con el perfil de un jugador que desempeña esa posición. Las variables que aportan positivamente a la calificación son el porcentaje de efectividad de los pases y los centros, las disputas defensivas ganadas, el porcentaje de disputas por arriba ganadas, el porcentaje de regates con éxito, la cantidad de aceleraciones entre 3 y 10 m/s² y la velocidad máxima (m/s); mientras que las variables que restan al puntaje son las tarjetas rojas, las faltas cometidas, el porcentaje de disputas en ataque ganadas, las disputas aéreas, las entradas efectivas y la distancia recorrida (Tabla 3).

Tabla 3. Variables del modelo de regresión lineal ajustado para defensas laterales.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	38.111	40.902	0.932	0.353
`Ocasiones generadas`	9.123	5.185	1.759	0.081
Faltas	-3.946	1.468	-2.688	0.008
`Tarjetas rojas`	-23.763	11.549	-2.058	0.042
`% de efectividad de pases`	103.591	29.172	3.551	0.001
Centros	2.621	1.202	2.18	0.031
`% de efectividad de los centros`	11.021	6.353	1.735	0.085
`Balones perdidos`	1.587	0.975	1.627	0.106
`Pérdidas en campo propio`	2.853	1.861	1.533	0.128
`Balones recuperados`	-1.256	0.691	-1.818	0.071
`Disputas defensivas ganadas`	5.895	1.217	4.844	0.001
`% disputas defensivas ganadas`	10.88	9.057	1.201	0.232
`% disputas en ataque ganadas`	-15.1	6.669	-2.264	0.025
`Disputas aéreas`	-4.316	1.123	-3.843	0.001
`% disputas por arriba ganadas`	15.622	5.087	3.071	0.003
`Regates con éxito, %`	17.981	4.813	3.736	0.001
`Robos de balón con éxito, %`	4.571	6.305	0.725	0.470
Interceptaciones	-1.168	0.746	-1.565	0.120
`Entradas efectivas`	-3.505	1.661	-2.11	0.037
`Pases de finalización efectivos`	12.148	6.428	1.89	0.061
`Acciones efectivas`	0.176	0.138	1.278	0.203
Resultado=Empate	3.305	4.844	0.682	0.496
Resultado=Victoria	20.261	4.388	4.617	0.001
`Distancia (m)`	-0.003	0.001	-2.382	0.019
`Velocidad Absoluta [12 - 19](m)`	0.005	0.004	1.249	0.214
`Velocidad Absoluta [19 - 24](m)`	-0.038	0.019	-1.962	0.052
`Acc Absoluta [3 - 10]`	0.405	0.183	2.206	0.029
`MAX Speed(m/s)`	1.899	0.949	2.001	0.048

Nota: Acc= Aceleraciones; Speed = Velocidad.

En el caso de los **mediocampistas defensivos**, se creó un modelo con 157 casos que logró obtener una capacidad predictiva de 70.9%. Se determinaron como variables significativas relevantes para la posición y además con coeficiente positivo las asistencias, oportunidades de gol con éxito, las ocasiones generadas, los tiros a portería, porcentaje de efectividad de los centros, el porcentaje de regates con éxito, las interceptaciones, las disputas

en ataque ganadas, entradas efectivas, porcentaje de acciones efectivas, los pases de finalización y la distancia recorrida (m). Las variables con coeficiente negativo fueron las pérdidas en campo propio, número de centros efectivos, tiros por fuera, los sprints absolutos y la condición de visitante (Tabla 4).

Tabla 4. Variables del modelo de regresión lineal ajustado para mediocampistas defensivos.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	119.151	23.529	5.064	0.001
Asistencias	10.168	4.946	2.056	0.042
`Jugadas de gol`	-7.613	4.463	-1.706	0.091
`Oportunidades de gol con éxito`	27.323	7.122	3.837	0.001
`Ocasiones generadas`	9.389	4.282	2.192	0.030
Tiros	9.962	2.380	4.186	0.001
`Tiros a portería`	-5.638	3.129	-1.802	0.074
`Pases de finalización`	-3.085	2.152	-1.434	0.154
`% de efectividad de los centros`	40.394	10.745	3.759	0.001
`Balones perdidos`	1.455	0.699	2.083	0.039
`Pérdidas en campo propio`	-5.841	1.471	-3.970	0.001
`Balones recuperados`	1.097	0.721	1.522	0.131
`% disputas ganadas`	13.727	9.306	1.475	0.143
`% disputas en ataque ganadas`	-6.483	5.036	-1.287	0.200
`Regates con éxito, %`	8.780	4.088	2.148	0.034
Intercepciones	2.371	0.705	3.362	0.001
Rechaces	-0.993	0.533	-1.863	0.065
`Disputas en ataque ganadas`	3.277	1.038	3.158	0.002
`Entradas efectivas`	3.928	1.228	3.199	0.002
`Pases de finalización efectivos`	10.643	5.157	2.064	0.041
`Centros efectivos`	-19.838	4.763	-4.165	0.001
`% acciones efectivas`	41.310	17.445	2.368	0.019
`Tiros fuera`	-7.357	3.006	-2.448	0.016
Resultado=Empate	-4.058	4.331	-0.937	0.351
Resultado=Victoria	12.901	3.761	3.431	0.001
Condición=Visita	-8.361	2.821	-2.964	0.004
`Distance(m)`	0.002	0.001	2.189	0.030
`Velocidad Absoluta [12 - 19](m)`	-0.007	0.004	-1.820	0.071
`Velocidad Absoluta [19 - 24](m)`	0.046	0.025	1.822	0.071
`Acc Abs [3-10]`	-0.242	0.141	-1.715	0.089
`Sprints ABS`	-1.545	0.534	-2.892	0.005
`MAX Speed(m/s)`	1.231	0.710	1.732	0.086

Nota: Acc= Aceleraciones; Speed = Velocidad.

Para la **posición de mediocampistas creativos ofensivos** se realizó un modelo con 168 casos, el cual presenta capacidad predictiva de 79.9%. Se determinaron como variables

significativas relevantes para la posición y además con coeficiente positivo las asistencias, el porcentaje de conversión de las ocasiones de gol, las jugadas de gol, los tiros a portería, el porcentaje de efectividad de los pases, las pérdidas en campo propio, las recuperaciones en campo rival, las disputas defensivas ganadas, las disputas por arriba ganadas, los regates efectivos, los pases de finalización efectivos, la asistencia esperada y las capacidades de desacelerar y de desarrollar velocidad máxima. En cuanto a las variables con coeficiente negativo, se encuentran: los goles esperados, los balones perdidos, las disputas defensivas, las disputas aéreas, el porcentaje de disputas por arriba ganadas, y el porcentaje de efectividad en lanzamientos (Tabla 5).

Tabla 5. Variables del modelo de regresión lineal ajustado para mediocampistas ofensivos.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	108.638	23.922	4.541	0.001
Asistencias	11.404	3.542	3.219	0.002
`Jugadas de gol`	9.797	2.304	4.253	0.001
`Ocasiones de gol, % conversión`	35.289	4.303	8.201	0.001
Faltas	2.654	1.367	1.942	0.054
`Tiros a portería`	17.216	2.655	6.485	0.001
`Goles esperados`	-20.396	6.973	-2.925	0.004
`% de efectividad de pases`	54.095	15.487	3.493	0.001
`Balones perdidos`	-2.679	0.581	-4.612	0.001
`Pérdidas en campo propio`	3.33	1.411	2.360	0.020
`Recuperaciones en campo rival`	2.934	1.428	2.055	0.042
`Disputas defensivas`	-2.259	0.894	-2.527	0.013
`% disputas en ataque ganadas`	9.776	5.565	1.757	0.081
`Disputas aéreas`	-4.275	1.306	-3.274	0.001
`% disputas por arriba ganadas`	-23.326	7.083	-3.293	0.001
`Disputas defensivas ganadas`	8.976	1.472	6.096	0.001
`Disputas por arriba ganadas`	15.462	3.667	4.217	0.001
`Regates efectivos`	5.597	0.968	5.783	0.001
`Pases de finalización efectivos`	5.78	2.282	2.533	0.012
`% de efectividad en lanzamientos`	-19.723	5.505	-3.583	0.001
`Asistencia esperada`	13.349	6.742	1.980	0.050
Resultado=Empate	-0.15	4.279	-0.035	0.972
Resultado=Victoria	12.233	3.669	3.334	0.001
`Distance(m)`	-0.002	0.001	-1.697	0.092
`Acc Abs [-10 -03 m/s]`	0.293	0.128	2.295	0.023
`MAX Speed(m/s)`	1.443	0.698	2.069	0.040

Nota: Acc= Aceleraciones; Speed = Velocidad.

Para los **extremos delanteros**, el modelo se ajustó basado en 201 casos y su capacidad predictiva es de 81.8%. Las variables significativas más relevantes para la posición y además con coeficiente positivo fueron las asistencias, los goles, las ocasiones generadas, los tiros a portería, el porcentaje de efectividad de pases, el porcentaje de disputas ganadas, regates, entradas efectivas, los centros, los tiros interceptados, las acciones totales, el asistencia esperada, mientras que las variables con coeficiente negativo fueron los balones perdidos, las disputas, los centros, las disputas en ataque ganadas, y la distancia medida en m/min (Tabla 6).

Tabla 6. Variables del modelo de regresión lineal ajustado para extremos delanteros.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	185.189	16.739	11.064	0.001
Goles	36.881	3.005	12.275	0.001
Asistencias	13.108	3.352	3.910	0.001
`Ocasiones generadas`	8.912	1.580	5.639	0.001
`Fuera de juego`	4.587	2.063	2.223	0.028
`Tiros a portería`	8.333	2.272	3.667	0.001
`% de efectividad de pases`	23.548	11.721	2.009	0.046
Centros	-1.639	0.804	-2.039	0.043
`Balones perdidos`	-1.006	0.461	-2.181	0.031
`% disputas ganadas`	44.143	10.078	4.380	0.001
Regates	2.847	0.678	4.197	0.001
`Regates con éxito, %`	7.877	4.181	1.884	0.061
Entradas	-1.478	1.017	-1.454	0.148
Rechaces	-1.255	0.635	-1.977	0.050
`Disputas en ataque ganadas`	-1.625	0.822	-1.975	0.050
`Entradas efectivas`	3.613	1.707	2.117	0.036
`Centros efectivos`	4.667	1.947	2.397	0.018
`Acciones totales`	0.364	0.090	4.033	0.001
`% de efectividad en lanzamientos`	-8.460	5.856	-1.445	0.150
`Tiros interceptados`	4.235	2.129	1.989	0.048
`Asistencias esperadas`	13.388	4.848	2.761	0.006
Resultado=Empate	3.035	3.491	0.869	0.386
Resultado=Victoria	12.789	3.227	3.964	0.001
Condición=Visita	-4.173	2.467	-1.692	0.092
`Dist (m/min)`	-0.383	0.123	-3.122	0.002

Se contó con 151 casos de **centros delanteros** que permitieron ajustar un modelo con una capacidad predictiva de 82.37%. Las variables que aportaron positivamente a la

calificación asignada a los jugadores que se desempeñan en esta posición fueron los goles, las asistencias, los tiros a portería, los pases, los pases de finalización, el porcentaje de efectividad de los centros, las diputadas, los pases efectivos de finalización, los tiros por fuera, y la cantidad de sprints realizados. En contraparte, las variables que afectan negativamente a la calificación son las tarjetas amarillas, los goles esperados, fueros de juego y los balones perdidos, los regates efectivos, los centros efectivos y la distancia recorrida a velocidades entre 12-19 km/h (Tabla 7).

Tabla 7. Variables del modelo de regresión lineal ajustado para centros delanteros.

Variable	Coeficiente	Error Estándar	Valor de Probabilidad Asociada	
			t tabular	P ≤
Intercepto	165.593	5.912	28.010	0.001
Goles	31.334	3.927	7.979	0.001
Asistencias `Ocasiones de gol, % conversión`	22.431	4.562	4.917	0.001
`Ocasiones generadas`	11.775	6.210	1.896	0.060
`Tarjetas amarillas`	2.590	1.906	1.359	0.177
`Tarjetas amarillas`	-10.325	3.311	-3.118	0.002
`Fueros de juego`	-3.560	1.259	-2.828	0.006
`Tiros a portería`	8.979	2.010	4.467	0.001
`Goles esperados`	-11.499	4.757	-2.417	0.017
Pases	0.782	0.157	4.995	0.001
`% de efectividad de pases`	6.216	5.543	1.121	0.264
`Pases de finalización`	6.964	2.269	3.069	0.003
`% de efectividad de los centros`	55.688	10.421	5.344	0.001
`Balones perdidos`	-1.545	0.606	-2.551	0.012
Disputas	0.922	0.386	2.388	0.018
`% disputas ganadas`	13.578	8.191	1.658	0.100
`% disputas defensivas ganadas`	-5.686	4.110	-1.383	0.169
Regates	1.800	1.381	1.303	0.195
`Regates con éxito, %`	7.629	4.571	1.669	0.098
`Robos de balón con éxito, %`	7.377	3.827	1.928	0.056
`Regates efectivos`	-6.057	2.750	-2.202	0.030
`Pases de finalización efectivos`	15.934	3.614	4.410	0.001
`Centros efectivos`	-55.236	8.093	-6.825	0.001
`Tiros fuera`	4.931	1.579	3.123	0.002
`Tiros interceptados`	3.965	2.599	1.526	0.130
`Asistencias esperadas`	-14.050	7.313	-1.921	0.057
Resultado=Empate	-9.867	3.465	-2.847	0.005
Resultado=Victoria	1.999	3.306	0.604	0.547
`Vel Abs [12 - 19](m)`	-0.008	0.002	-4.167	0.001
`Sprints ABS`	0.618	0.222	2.788	0.006

c. Modelos de regresión logística por posición en el campo de juego

Para los **defensas centrales**, la cantidad de tiros realizados, el porcentaje de efectividad de los lanzamientos, la distancia recorrida en metros y los metros recorridos por minuto influyeron positivamente en la probabilidad de ganar un partido, mientras que las faltas cometidas, las tarjetas rojas, los tiros a portería, los goles esperados, los regates efectivos, los pases efectivos y la distancia recorrida en alta velocidad disminuyeron la probabilidad de tener un resultado favorable en el partido (Tabla 8). El modelo ajustado para esta posición tuvo una capacidad de acierto de 69.2%.

Tabla 8. Variables significativas para el modelo de regresión logística para defensas centrales.

Variable	Coefficiente
Intercepto	-2.267
Faltas	-0.183
`Tarjetas rojas`	-16.257
Tiros	0.714
`Tiros a portería`	-3.653
`Goles esperados`	-5.497
`Regates efectivos`	-0.416
`Pases efectivos`	-0.016
`% de efectividad en lanzamientos`	4.024
`Distance (m)`	0.0002
`HSR Abs dist (m)`	-0.004
`Dist (m/min)`	0.033

Con respecto a los **defensas laterales**, se identificaron como variables que aumentaron la probabilidad de obtener una victoria a los goles, la cantidad de ocasiones generadas, las tarjetas rojas, las recuperaciones en campo rival, la distancia recorrida a una velocidad entre 19 y 24 km/h, los metros recorridos por minuto y la cantidad de aceleraciones entre 3 y 10 m. Las variables que disminuyeron las probabilidades de ganar fueron las tarjetas amarillas, los fueros de juego, los balones perdidos, las disputas en ataque ganadas, el porcentaje de acciones efectivas, las asistencias esperadas y la distancia recorrida en alta velocidad (HSR) (Tabla 9). Para esta posición, el modelo ajustado tuvo una capacidad de acierto de 68.3%.

Tabla 9. Variables significativas para el modelo de regresión logística para defensas laterales.

Variable	Coefficiente
Intercepto	0.562
Goles	16.573
`Ocasiones generadas`	1.013
`Tarjetas amarillas`	-0.988
`Tarjetas rojas`	17.401
`Fuera de juego`	-0.611
`Balones perdidos`	-0.226
`Recuperaciones en campo rival`	0.248
`Disputas en ataque ganadas`	-0.286
`% acciones efectivas`	-5.182
`Asistencias esperadas`	-3.185
`HSR Abs dist (m)`	-0.002
`Vel Abs [19 - 24](m)`	0.002
`Dist (m/min)`	0.027
`Acc Abs [3- 10]`	0.055

En los **mediocampistas defensivos**, se identificó que los goles, las asistencias, las ocasiones generadas, los goles esperados, el porcentaje de disputas ganadas, las interceptaciones de balón, los regates efectivos, el porcentaje de efectividad en lanzamientos, los metros recorridos por minuto y las aceleraciones realizadas entre 3 y 10 m contribuyeron a la probabilidad de ganar un partido. Por el contrario, redujeron la posibilidad de ganar las jugadas de gol donde participaron, los fueros de juego cometidos, los tiros a portería, el porcentaje de regates con éxito, el porcentaje de robos de balón realizados con éxito, las disputas en ataque ganadas, el porcentaje de acciones efectivas, las asistencias esperadas y la distancia recorrida en alta velocidad (Tabla 10). El modelo ajustado tuvo una capacidad de acierto del 73.9%.

Tabla 10. Variables significativas para el modelo de regresión logística para mediocampistas defensivos.

Variable	Coefficiente
Intercepto	0.676
Goles	3.860
Asistencias	2.380
`Jugadas de gol`	-1.255
`Ocasiones generadas`	1.236
`Fueras de juego`	-4.358
`Tiros a portería`	-1.544
`Goles esperados`	4.717
`% disputas ganadas`	2.802
`Regates con éxito, %`	-1.971
`Robos de balón con éxito, %`	-1.261
Interceptaciones	0.358
`Disputas en ataque ganadas`	-0.350
`Regates efectivos`	0.684
`% acciones efectivas`	-6.406
`% de efectividad en lanzamientos`	3.144
`Asistencias esperadas`	-4.413
`HSR Abs Dist (m)`	-0.005
`Dist (m/min)`	0.038
`Acc Abs [3 -10]`	0.027

Para los **mediocampistas ofensivos** se determinó que las variables favorables para ganar un partido fueron los goles, las asistencias, los tiros, las disputas defensivas ganadas, la velocidad registrada entre 19 y 24 m y la máxima velocidad; mientras que las variables que reducen las probabilidades de ganar fueron los tiros a portería, los balones perdidos, los robos de balón con éxito, la velocidad registrada entre 12 y 19 m y la cantidad de sprints realizados (Tabla 11). Para esta posición, el modelo tuvo una capacidad de acierto de 69.6%.

Tabla 11. Variables significativas para el modelo de regresión logística para mediocampistas ofensivos.

Variable	Coefficiente
Intercepto	-4.951
Goles	1.776
Asistencias	1.100
Tiros	0.363
`Tiros a portería`	-0.514
`Balones perdidos`	-0.131
`Robos de balón con éxito, %`	-0.928
`Disputas defensivas ganadas`	0.311
`Vel Abs [12 - 19](m)`	-0.001
`Vel Abs [19 - 24](m)`	0.005
`Sprints ABS`	-0.083
`MAX Speed(m/s)`	0.188

Para los **extremos delanteros**, las variables positivas que aumentaron la probabilidad de ganar un partido fueron las asistencias, las ocasiones generadas, las faltas recibidas, los goles esperados, las disputas defensivas ganadas, el porcentaje de acciones efectivas, el porcentaje de efectividad en los lanzamientos, los tiros fuera de la portería, la distancia recorrida en alta velocidad, la velocidad registrada entre 19 y 24 m, los metros recorridos por minuto y la máxima velocidad registrada. Por el contrario, las variables que disminuyeron la probabilidad de ganar fueron los goles anotados, los tiros realizados, el porcentaje de disputas ganadas, el porcentaje de regates con éxito, los pases efectivos y las desaceleraciones entre 3 y 10 m (Tabla 12). Este modelo tuvo una capacidad de acierto de 75.1%.

Tabla 12. Variables significativas para el modelo de regresión logística para extremos delanteros.

Variable	Coefficiente
Intercepto	-17.949
Goles	-0.933
Asistencias	0.837
`Ocasiones generadas`	0.561
`Faltas recibidas`	0.303
Tiros	-0.503
`Goles esperados`	2.730
`% disputas ganadas`	-6.461
`Regates con éxito, %`	-1.010
`Disputas defensivas ganadas`	0.258
`Pases efectivos`	-0.049
`% acciones efectivas`	9.384
`% de efectividad en lanzamientos`	2.026
`Tiros fuera`	0.691
`HSR Abs Dist (m)`	0.002
`Vel Abs [19 - 24](m)`	0.003
`Dist (m/min)`	0.054
`Acc Abs [-10 -03]`	-0.042
`MAX Speed(m/s)`	0.253

Finalmente, las variables que contribuyeron positivamente a ganar un partido para los **delanteros centro** fueron los goles anotados, las jugadas de gol creadas, las interceptaciones de pases, los regates efectivos, el porcentaje de efectividad en los lanzamientos y la distancia recorrida en alta velocidad. Las variables que redujeron la probabilidad de ganar fueron las faltas cometidas, así como las recibidas, los tiros a portería, el porcentaje de disputas aéreas ganadas, los pases efectivos, la distancia recorrida entre 12 y 19 m/s y la cantidad de desaceleraciones absolutas entre -10 y -3 m/s² (Tabla 13). Para este modelo se obtuvo una capacidad de acierto de 72.2%.

Tabla 13. Variables significativas para el modelo de regresión logística para delanteros centro.

Variable	Coefficiente
Intercepto	0.129
Goles	1.368
`Jugadas de gol`	0.552
Faltas	-0.353
`Faltas recibidas`	-0.367
`Tiros a portería`	-1.711
`% disputas por arriba ganadas`	-1.273
Interceptaciones	0.843
`Regates efectivos`	0.697
`Pases efectivos`	-0.120
`% de efectividad en lanzamientos`	2.006
`HSR Abs dist (m)`	0.007
`Vel Abs [12 - 19](m)`	-0.001
`Acc Abs [01- 03]`	-0.039

d. Modelos de árboles de decisiones

Se realizaron modelos de árboles de decisiones en los que se tunearon los parámetros de *min_n*, *cost_complexity* y finalmente *tree_depth*. Para cada una de las posiciones se puede ver los resultados de la precisión y el valor del área bajo la curva (AUC, **Area Under the Curve**) (Tabla 14).

Tabla 14. Indicadores del modelo de árboles de decisiones.

Posición	Precisión	AUC
Defensas	0.43	0.49
Delanteros	0.46	0.47
Mediocampistas	0.66	0.70
Contenciones	0.46	0.47
Extremos	0.48	0.49
Laterales	0.69	0.70

e. Modelos de bosques aleatorios

Seguidamente se realizaron los modelos de bosques aleatorios, para los cuales se tunearon los parámetros de *mtry*, *min_n* y se dejó la variable *trees* fija en 1000. Para cada una de las posiciones se puede ver los resultados de la precisión y el valor del AUC (Tabla 15).

Tabla 15. Indicadores del modelo de bosques aleatorios.

Posición	Precisión	AUC
Defensas	0.63	0.69
Delanteros	0.48	0.5
Mediocampistas	0.67	0.71
Contenciones	0.47	0.53
Extremos	0.53	0.54
Laterales	0.68	0.71

Como se observa en las tablas 14 y 15, el método que obtuvo mejores resultados en general es el de bosques aleatorios, ya que presenta mejores puntuaciones (más cercanas a 1) tanto en la precisión como en el AUC con respecto al modelo de árboles de decisión. En las figuras 10, 11, 12, 13, 14, 15 se muestran las variables de mayor importancia que logró verificar el modelo para cada una de las posiciones. Entre más puntaje tenga, más significativo es el indicador en el algoritmo y en su mayoría, fueron las variables recolectadas con el dispositivo WIMU PRO®; es decir, variables relacionadas con el rendimiento físico del futbolista en el partido.

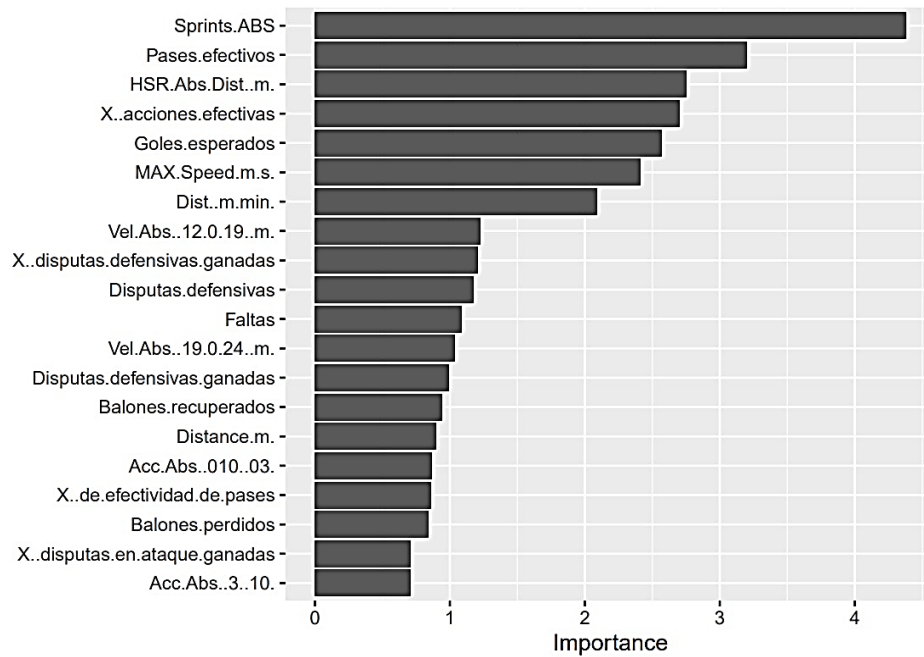


Figura 10. Variables más importantes para los defensas centrales.

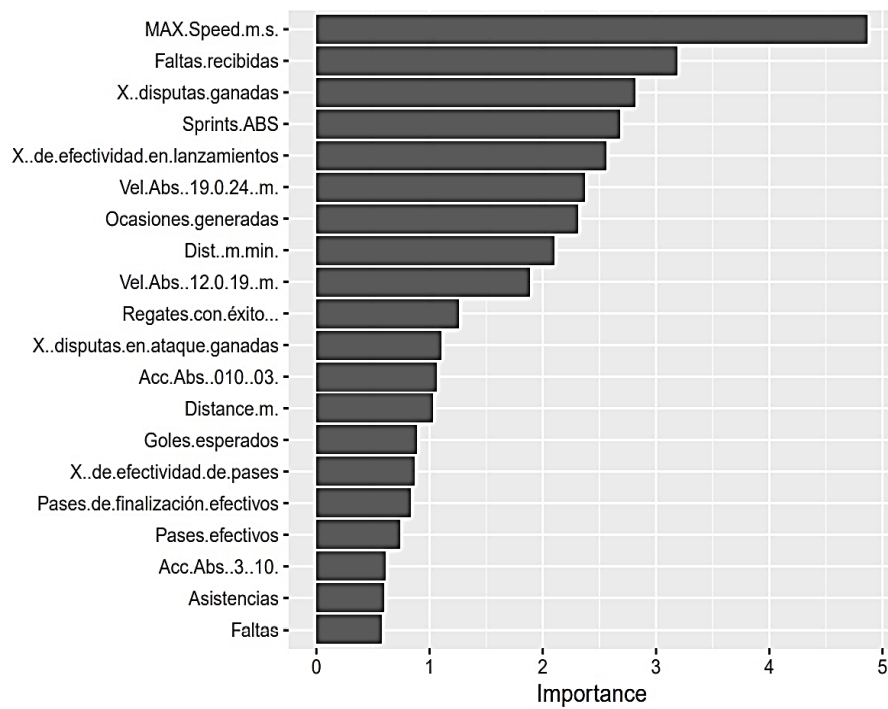


Figura 11. Variables más importantes para los defensas laterales.

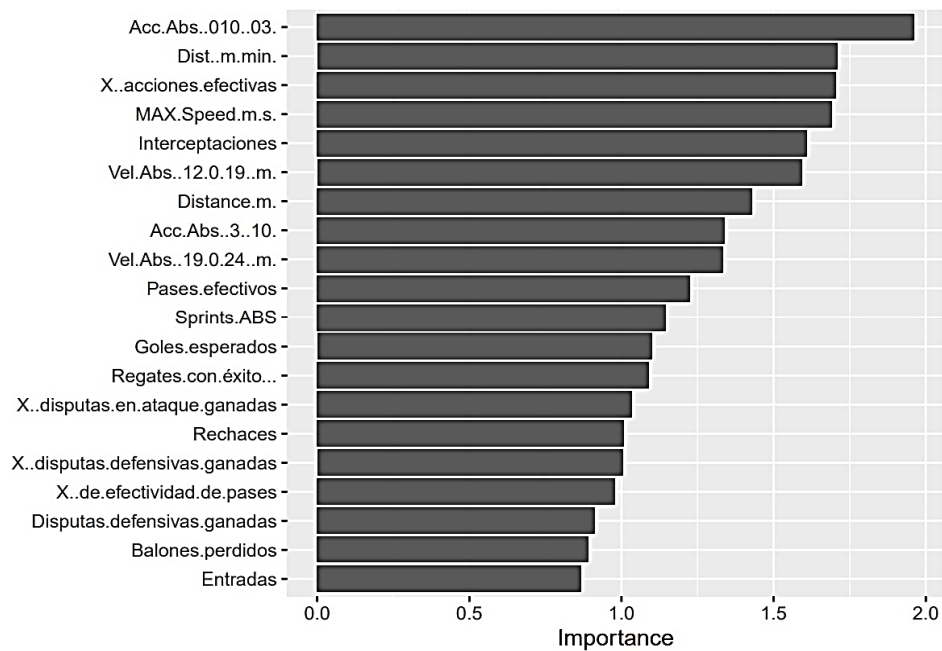


Figura 12. Variables más importantes para los mediocampistas defensivos.

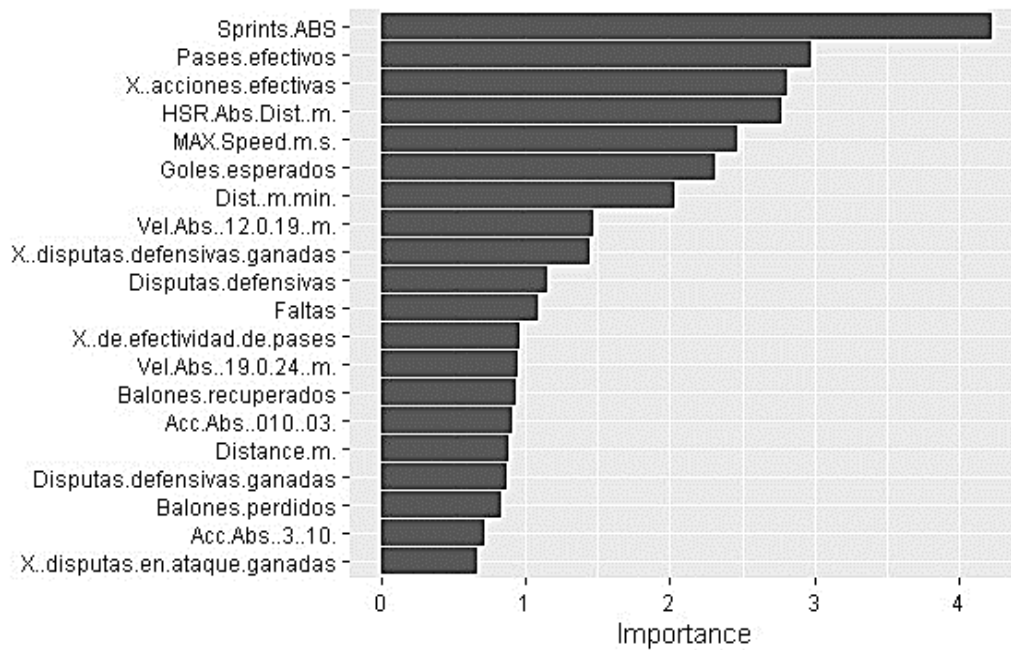


Figura 13. Variables más importantes para los mediocampistas ofensivos.

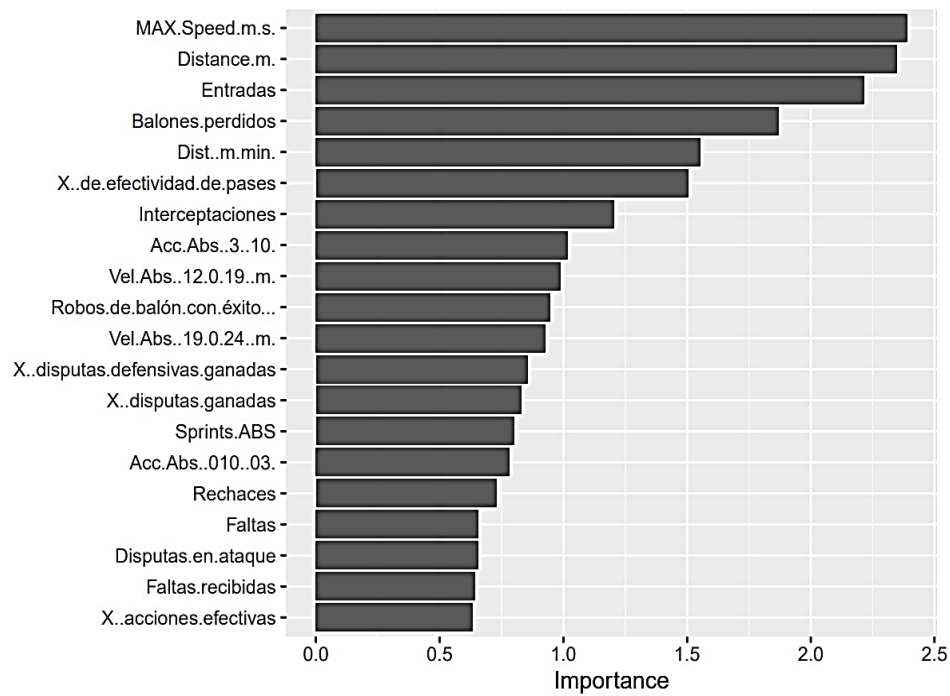


Figura 14. Variables más importantes para los extremos delanteros.

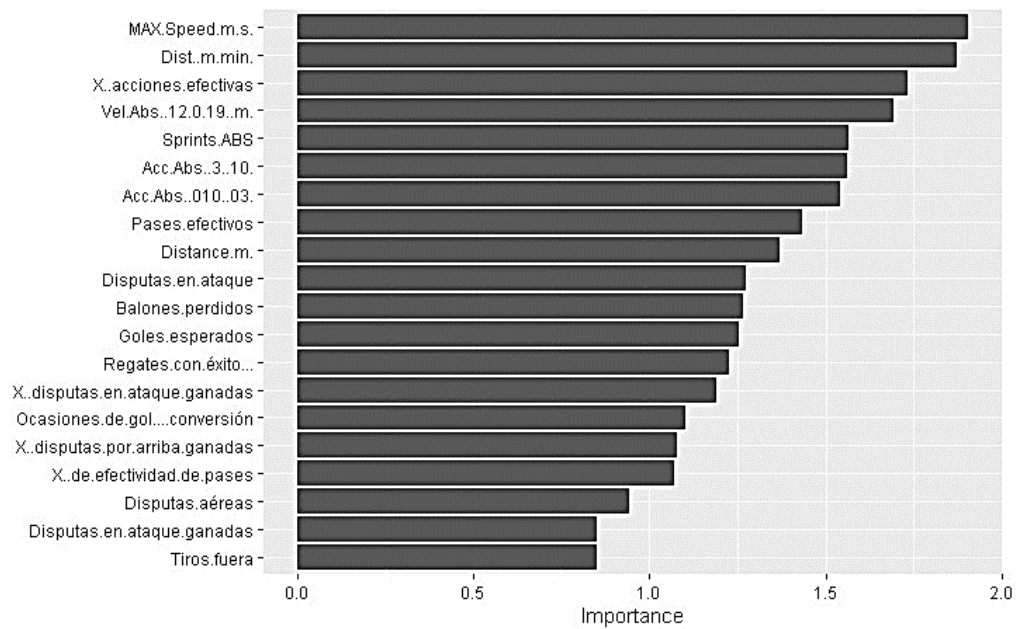


Figura 15. Variables más importantes para los centro delanteros.

Capítulo VI. Discusión

El objetivo general de esta tesis fue analizar retrospectivamente el rendimiento de los jugadores, por posición, de un equipo de fútbol profesional costarricense a partir de variables físicas y técnico-tácticas recolectadas por medio de tecnología inercial WIMU PRO® y de la plataforma de datos Instat®.

Como punto de partida y para poder discutir los resultados es importante describir el estilo de juego del equipo evaluado, pues como se ha descrito anteriormente, este es un factor vital que condiciona el comportamiento y la estrategia durante el partido (Lago-Peñas et al., 2017). Conocer este detalle, permite analizar los resultados en un contexto más acorde y con una interpretación aún más clara. También es importante anotar que durante el período de tiempo estudiado el equipo logró mantenerse siempre en las primeras tres posiciones en cuatro ediciones del Campeonato Promérica de Primera División. Además, consiguió un título centroamericano en el Torneo de Liga CONCACAF Scotiabank.

El sistema más frecuentemente utilizado fue el 1-4-3-3, en donde los jugadores más ofensivos se ubicaron en el campo rival; por ejemplo, el centro delantero dando profundidad y los extremos abiertos pegados a las bandas. Los laterales ubicados casi a mitad de campo y el mediocampista defensivo generando las salidas junto a los dos centrales. El sistema ofensivo se caracterizó por la construcción del juego desde la salida con el balón dominado, y la combinación con paciencia, buscando posesiones largas del balón y generando los ataques de forma posicional. El porcentaje de posesión de balón durante los partidos promedió entre un 50% y 60% más alto que el del rival, lo que lo ubicó siempre en el primero o segundo lugar de los equipos con más posesión del torneo. También fue utilizado en algunos momentos el juego directo, buscando ser muy intensos en las segundas jugadas (i.e., rebotes) para recuperar y acelerar el juego. Por otra parte, en fase defensiva, se intentaba intensificar la presión una vez perdida la pelota y si no era exitosa en los primeros segundos, se replegaba por completo en zona media.

6.1. Identificación de variables que explican el rendimiento de un futbolista

Con el fin de determinar las variables que explican el rendimiento de un futbolista en un juego de fútbol, a partir del índice de rendimiento o Índice Instat®, se elaboraron seis modelos de regresión lineal que lograron otorgar una calificación a los jugadores que

participaron en los partidos de acuerdo a las variables medidas durante el juego y que fueron consideradas estadística y futbolísticamente importantes para cada posición táctica.

El principal hallazgo, desde el punto de vista estadístico, fue que todos los modelos mostraron una adecuada capacidad predictiva del rendimiento para las diferentes posiciones de los jugadores en el terreno de juego. Así, se obtuvieron valores de 0.67 para defensas centrales, 0.56 para defensas laterales, 0.71 para los mediocampistas defensivos, 0.82 para los extremos, de 0.80 para los mediocampistas creativos, y 0.82 para los delanteros centro. Además, es importante destacar que todos los modelos cumplieron con los supuestos de normalidad, homocedasticidad, linealidad y no multicolinealidad.

Otro aspecto que se debe resaltar es que se consiguió ajustar modelos elaborados con variables que cambian según la posición. El objetivo de crear seis modelos distintos se basaba en la necesidad de diferenciar qué características o aspectos eran influyentes para determinar si una actuación durante el partido es buena o mala de acuerdo a los diferentes roles que deben cumplir los jugadores dentro de la cancha. Por ejemplo, mientras el porcentaje de disputas ganadas resultó ser importante para los defensas, para los extremos fueron los regates, y para un mediocampista creativo, fueron las ocasiones de gol creadas durante el partido.

6.1.1. Defensas centrales

Con un modelo predictivo del 67%, uno de los factores que discriminaron el rendimiento de los jugadores de fútbol fue el porcentaje de acciones efectivas, asociado a la eficacia con la que este jugador logró ejecutar todas sus acciones técnicas con la mayor destreza posible. Entre estas acciones destacan las ofensivas como los tiros, las asistencias, los goles, los pases efectivos, y los pases de finalización efectivos.

El aporte que estos jugadores ofrecen a nivel ofensivo en cuanto a tiros, asistencias y goles, ha sido manifestado en varios estudios que lo destacan como una faceta fundamental del defensor para un equipo exitoso (Ade et al., 2016; Barron et al., 2020; Bradley & Ade, 2018). La pelota quieta ofensiva les brinda la oportunidad de sumarse al ataque y tener posibilidades de anotar (Ade et al., 2016). Cabe destacar que, el somatotipo que presentaron estos jugadores (i.e., los más altos del equipo) fue clave para ganar muchos balones aéreos en estas acciones. Posiblemente por esta situación, el sumar muchos tiros por fuera le restaron

al defensor central en la calificación en el modelo logrado por lo que para efectos de mejora será una variable considerable a la hora de ejecutar los programas de entrenamiento.

Para Barron et al. (2020) un defensor central deberá tener una buena precisión de tiro, un promedio de tiro alto y un promedio de asistencia alto, para lograr alcanzar otro nivel profesional. Los autores lograron crear un modelo con una predicción perfecta (100%), en la que estas variables tuvieron un gran impacto en la discriminación de jugadores que habían actuado en una categoría inferior y al siguiente año lograron participar en la Premier League inglesa.

Como se mencionó en el capítulo del marco teórico, el pase es una cualidad que ha evolucionado en los últimos años en esta posición ya que el fútbol se dirige cada vez más hacia un juego de combinación, por lo que, tener la capacidad de pasar el balón con precisión a distintas distancias y direcciones es un factor clave para el rendimiento (Adams et al., 2013; Barron et al., 2020; Bradley et al., 2013).

En este caso la gran precisión en el pase que mostraron los jugadores en esta posición se explica porque los jugadores medidos en el estudio, debían tener la facilidad de “sacar el balón” o salir jugando desde atrás a través del pase. Si bien es cierto muchas veces lograron asociarse con jugadores en posiciones más ofensivas (e.g., mediocampistas), hubo momentos del partido, en los cuales para romper líneas de presión del rival intentaron buscar el lanzamiento como alternativa. Esta acción de pase en sí, no parece estar orientada hacia un objetivo específico, por lo que, para efectos del algoritmo en el índice de rendimiento, un lanzamiento del defensor puede relacionarse a un pobre rendimiento técnico, a pesar de ser efectivo, ya que esta se asocia más al juego directo.

Desde el punto de vista defensivo, las variables más significativas fueron las entradas efectivas y el porcentaje de disputas defensivas ganadas. Estas son acciones vitales en el desempeño del defensor central (Konefał, Chmura, Kowalczyk, et al., 2019; Modric et al., 2022), quien debe velar continuamente por el marcaje, reconociendo por ejemplo en la acción de entrada, cuándo o cómo salir por el adversario ya que esto lo obliga a perder su posición.

Según Modric et al. (2019) tanto las entradas, como los duelos aéreos, y las jugadas a balón parado en defensa han sido tomadas en cuenta a la hora de calcular el Índice Instat[®], por lo que existe una estrecha relación con los hallazgos encontrados.

En el presente estudio ningún indicador físico logró explicar el rendimiento de los defensores centrales a partir del índice de rendimiento.

6.1.2. Defensas laterales

El modelo creado para esta posición fue el de menor capacidad predictiva, con apenas un 56%. Sin embargo, las variables que se obtuvieron como significativas definen muy bien el papel que desempeña un defensa lateral, por ejemplo, el porcentaje de efectividad en los pases, los centros, las disputas defensivas ganadas, el porcentaje de disputas aéreas ganadas, el porcentaje de regates con éxito, la velocidad máxima y la aceleración (3-10 m/s).

Algunos de estos parámetros técnicos son utilizados para la elaboración del índice de rendimiento utilizado (Modric et al., 2022), por ejemplo, en la figura 16, se aprecia el cálculo para un defensor lateral, encontrándose factores importantes como: la precisión en los pases, los centros, el porcentaje de regates con éxito, porcentaje de duelos ganados, etc.

Parámetros técnicos	Cantidad/Porcentaje	Factor
Precisión del pase	83.60%	1.075104
Pases claves y pases adicionales de ataque	1	1.075104
Pases precisos en el área	1	1.024528
Tiros (cantidad)	0	1.007670
Cantidad de conducción de balón exitosas	23.10%	1.041387
Participación en acción de gol	25	1.053333
Errores graves	0	1.007670
Pases claves imprecisos y pases adicionales de ataque	1	1.016667
Cantidad de disputas ganadas	90.10%	1.142537
Tiros fuera del marco	1	1.016667
	Multiplicación de todos los factores =	1.557639
	Coeficiente ponderado del juego que considera el nivel del jugador y el nivel de sus compañeros de equipo y de los jugadores rivales =	0.998817
	Nivel base del índice InStat =	220
	Índice InStat final =	$1.557639 \times 0.9988165276 \times 220 = 342$

** No se incluyen todos los parámetros técnicos

Figura 16. Ejemplo del cálculo del Índice InStat® para un defensor lateral

De acuerdo a la filosofía de juego del equipo estudiado descrita anteriormente, los laterales tenían libertad para subir a atacar, generalmente a través de la combinación de pases (e.g., mediocampista ofensivo o extremo delantero), o también en situaciones en las que el equipo lograba dar cambios de juego y el callejón estaba “abierto” para ellos.

Estos jugadores debían interpretar cuándo debían jugar por las bandas para dar amplitud o cuándo jugar por dentro; es decir, hacia el centro del campo, para poder combinar y llevar el balón hacia el frente. Esto les generó estar continuamente en contacto con el balón, y sumar al puntaje de acuerdo a la efectividad con la que transfería el mismo. Adams et al. (2013), describieron la gran influencia que tienen los defensas laterales sobre las secuencias de pase en todas las zonas del campo.

También las subidas al ataque abrieron la posibilidad de producir muchos enfrentamientos ofensivos de 1 vs. 1, o ejecutar centros al área rival, acciones que son catalogadas como fundamentales para los jugadores que participan en esta posición y han sido relacionadas con el índice de rendimiento (Ade et al., 2016; Modric et al., 2019).

En cuanto a los indicadores claves defensivos, se han encontrado las entradas y las disputas aéreas (Ade et al., 2016; Hughes et al., 2012). Las entradas son acciones que se caracterizan por ser ejecutadas después de esfuerzos de alta intensidad (Ade et al., 2016), por la necesidad de acercarse y acechar al rival para arrebatar el balón. Los defensas laterales suelen preceder los esfuerzos con un giro de 90° a 180° en su transición de funciones ofensivas a defensivas, y ejecutan más entradas tras el esfuerzo que otras posiciones (Bradley & Ade, 2018). Contrariamente en el presente estudio, entre más frecuentes eran estas acciones, menor rendimiento del lateral, sugiriendo que el éxito futbolístico en el índice de rendimiento está más relacionado por sus acciones ofensivas, independientemente del hecho de que sean defensas (Modric et al., 2019).

Para las variables físicas, la carrera a máxima velocidad y las aceleraciones, fueron variables significativas que aportaron de forma positiva dentro del modelo. La eficacia en el juego por los laterales ha estado siempre asociada a la distancia recorrida a alta intensidad (> 21 km/h), así como al número de sprints (> 24 km/h) realizados (Andrzejewski et al., 2018; Chmura et al., 2018). La carrera a máxima velocidad se manifiesta en el cumplimiento del rol, sobre todo ofensivo, cuando este jugador se proyecta constantemente al ataque (Konefał, Chmura, Kowalczyk, et al., 2019; Modric et al., 2019; Oliva-Lozano et al., 2020). La

realización de estas acciones requiere como elemento clave la capacidad de acelerar y repetir esta acción continuamente (Akenhead et al., 2016; Harper et al., 2019).

Algunos indicadores significativos restaron explicación de varianza al modelo; por ejemplo, la distancia en las carreras a alta intensidad (19-24 km/h) y la distancia total, que quiere decir, que a mayor distancia recorrida menor rendimiento. Sin embargo, la literatura afirma que altos puntajes en la carrera a alta intensidad es una variable predictora para ganar en los partidos (Andrzejewski et al., 2018; Bradley & Ade, 2018). Estos jugadores deben estar en posiciones defensivas muy específicas cuando no se tiene el balón (e.g., formando la línea de 4), pero deben tener atributos ofensivos significativos, los cuales fueron descritos anteriormente. Esta “doble” función implicaría un constante desplazamiento de un lugar a otro, sumando mucha distancia y a una intensidad muy alta. Posiblemente estos esfuerzos fueron más orientados a acciones defensivas (e.g., coberturas, carreras de recuperación) que se realizan a alta intensidad (Bradley & Ade, 2018), interpretándose como una debilidad según el algoritmo creado para evaluar el rendimiento.

Por otra parte, en cuanto a las medidas disciplinarias, las tarjetas rojas y las faltas restaron al rendimiento de los defensas laterales. Sin embargo, estas variables no pueden ser controladas por los jugadores, pues dependen de la decisión arbitral, la cual es de naturaleza subjetiva.

6.1.3. Mediocampistas defensivos

El modelo creado tiene una capacidad predictiva alta, del 71%. Entre las variables ofensivas que aportan a aumentar la nota o el rendimiento del mediocampista defensivo se encontraron las oportunidades de gol con éxito, las asistencias, las ocasiones generadas, el % de efectividad de los centros, el porcentaje de efectividad de los regates, los tiros, los pases de finalización efectivos y el porcentaje de acciones efectivas.

El porcentaje de acciones efectivas, como se mencionó anteriormente, es un indicador de eficacia de todas las acciones que realiza el jugador durante el partido. Tiene un alto coeficiente estadístico que se ve reflejado por la gran cantidad de variables significativas positivas.

Los mediocampistas defensivos tenían un papel preponderante en la construcción del juego desde atrás, eran jugadores con muy buena destreza técnica, pasadores y con una gran capacidad de remate. No se caracterizaban por ingresar al área rival cuando el equipo iba al

frente y generalmente se ubicaban en una posición centralizada en la cancha con el marco rival relativamente cerca y de frente (30-40 metros de distancia) que les permitía fungir como apoyo, distribuidores de juego o incluso rematadores de media y larga distancia. Estos factores como la distancia, el origen de la acción, la calidad del jugador, aunado a la zona del cuerpo con la que se realiza el remate y la calidad del portero rival, se toman en consideración para calcular la oportunidad de gol (Eggels et al., 2016).

En cuanto al porcentaje de efectividad en los centros y las asistencias a gol, se debe a la participación activa en los cobros de pelota quieta como tiros de esquina o tiros libres, acciones que, de acuerdo a literatura previa, promueven entre el 30% y el 40% de los goles marcados en la élite (Lago-Peñas et al., 2017).

Por otra parte, la pérdida de balón en campo propio, calificó negativamente y fue recurrente, por la dependencia del equipo en este jugador para sacar el balón desde atrás con el fin de entregar el “balón limpio”; es decir, lo más claro posible para los jugadores más ofensivos del equipo. La pérdida continua de balones en campo propio comprometía al equipo que al estar en una posición ofensiva (“larga o abierta” para aprovechar el campo y generar espacios en la defensa rival), muchas veces se vio sorprendido y con altas posibilidades de encajar una anotación en contra por la cercanía con el propio marco. Sin embargo, este era un “riesgo calculado” por el tipo de planteamiento táctico del entrenador.

Desde el punto de vista defensivo las variables que discriminaron para tener un buen índice de rendimiento fueron las disputas ganadas en ofensiva, las entradas defensivas y las interceptaciones, que son acciones propias de la posición en la que se encuentra ya que requieren una habilidad para robar balones (Bradley et al., 2013).

El rendimiento de los jugadores que participaron en el estudio cuando jugaron de visita le restó a su calificación; es decir, fue un indicador negativo en la nota. Se puede argumentar que los mediocampistas defensivos, debido a los requisitos técnicos y tácticos específicos de su posición para dirigir el partido, se ven más afectados por el juego en casa que los centrocampistas anchos y los defensas. Esto puede deberse al hecho de que estos jugadores se caracterizan por mantener la posesión del balón de manera más prolongada y con un mayor número de comportamientos ofensivos (e.g., disparos, regates y toques de balón) cuando juegan en casa (Konefał, Chmura, et al., 2019a). Según Konefał et al. (2020),

las probabilidades de ganar en casa son diferentes para cada posición de juego; sin embargo, lograron estimar un 91.34% de posibilidades de ganar para los mediocampistas centrales.

Entre las variables físicas asociadas al índice Instat[®], se reporta la distancia recorrida, que mejora el puntaje y la cantidad de veces que alcanzó el sprint (> 24 km/h). Para la posición de mediocampistas ofensivos y defensivos, Modric et al. (2019) no lograron encontrar correlación de algún indicador físico con el índice de rendimiento y lo atribuyen más a una mayor demanda de los recursos técnicos-tácticos como la posesión del balón, el número de pases clave, los regates y los disparos para esta posición.

6.1.4. Mediocampistas ofensivos

Para los mediocampistas ofensivos, se diseñó un modelo con una capacidad predictiva alta del 79.9%, y las variables que explican el rendimiento fueron las asistencias, las asistencias esperadas, las jugadas de gol, las ocasiones de gol y el % de conversión, los tiros a portería, el porcentaje de efectividad de pases, regates efectivos y los pases de finalización efectivos.

Los mediocampistas ofensivos son creadores del juego ofensivo y se caracterizan por ejecutar pases claves dentro del juego (Modric et al., 2019). A diferencia de los mediocampistas defensivos, y por su ubicación más próxima al marco rival, tenían una mayor probabilidad de crear situaciones de gol. Debían “pisar” o entrar al área rival buscando la combinación, atacando la línea de los defensores o finalizando los centros enviados, lo que aumentaba el número de ocasiones de gol y tiros a portería. La identificación de algunas de estas variables fueron claves en el estudio de Barron et al. (2020), para la posición de mediocampistas ofensivos. Por ejemplo, el éxito del pase, las entradas al área de penal, los goles y los tiros a puerta, fueron elementos diferenciadores entre jugadores de diferente nivel competitivo.

Por otra parte, entre los resultados del presente trabajo, una mayor efectividad en los lanzamientos y perder muchos balones afectó negativamente la eficacia en el juego real para esa posición de juego. Esta situación, se puede explicar por un mal comportamiento táctico, ya que usualmente los mediocampistas ofensivos descendían para arrebatar el balón a los defensores. Esta insistencia por acarrear balones desde atrás, desorganizaba la estructura ofensiva y se perdían las referencias posicionales dentro de la cancha. Esto de alguna manera, aumentó el riesgo de perder el balón, o promovió el lanzamiento como una oportunidad de ir

al ataque, que si bien es cierto fue efectivo en cuanto a la acción en sí misma, era poco eficiente como forma de ataque por la poca cantidad de jugadores ubicados en esta zona.

Defensivamente las recuperaciones en campo rival, las disputas defensivas ganadas y disputas por arriba, fueron significativas y explican el índice de rendimiento. La presión tras pérdida, o sea, la intención de querer recuperar rápido el balón después de perderlo, fue una tarea que constantemente se logró ejecutar durante los partidos, sobre todo en campo rival. En cuanto a las disputas aéreas, la gran cantidad de pelotazos o rechaces aéreos del rival permitieron ser luchados sobre todo en el medio campo. El cabeceo, tanto defensivo como ofensivo es, según Hughes et al. (2012), un indicador importante de rendimiento para esta posición.

Desde el plano físico, las desaceleraciones y la velocidad máxima lograron ser acciones físicas que explicaron la varianza, aunque con coeficientes muy bajos.

6.1.5. Delanteros extremos

El modelo tuvo una capacidad predictiva bastante alta del 81.8% y algunas acciones técnicas ofensivas correlacionaron con el índice Instat[®]; por ejemplo, los goles, las asistencias, las ocasiones generadas, los tiros a portería y el porcentaje de efectividad de los pases, los centros efectivos, los regates. Tanto el regate como la finalización, son factores claves en el cálculo del índice de rendimiento (Modric et al., 2019).

Estos jugadores se caracterizaron por ser los más pequeños, con mucha explosividad y una gran riqueza técnica, lo que les permitía tener mucha capacidad en el dribbling. Su posición consistía en estar prácticamente pegados a la banda con la finalidad de “estirar” la línea defensiva contraria, aislando así a su marcador más próximo (defensa lateral) del defensor central. La finalidad era encarar en 1 vs. 1 para desbordar, centrar o recortar hacia adentro buscando rematar, aspectos que fueron significativos en el análisis creado en este estudio y que también han sido determinantes en otras investigaciones para el éxito de los equipos (Geurkink et al., 2021; Lago-Peñas et al., 2011; Taylor et al., 2008). Una abundante cantidad de centros, sin embargo, a pesar de ser un objetivo en la función del extremo, restó a la calificación, y esto puede justificarse a que los centros disminuyen cuando se gana y aumentan cuando se pierde (Fernandez-Navarro et al., 2018; Taylor et al., 2008). Desde esta perspectiva, el resultado final del partido tiene una estrecha relación con el índice Instat[®] para esta posición (Modric et al., 2022).

Defensivamente, las variables que le aportaron al modelo fueron las entradas efectivas y el porcentaje de disputas ganadas. Al igual que con los mediocampistas ofensivos, la capacidad de robar balones a través de la presión al contrincante elevó significativamente el rendimiento, comprobando lo que ha sido reportado en trabajos anteriores (Almeida et al., 2014; Modric et al., 2019; Vogelbein et al., 2014).

Desde el punto de vista físico, cuando el extremo delantero aumentó su distancia por tiempo (i.e., m/min) durante el partido la probabilidad de ganar disminuyó, y esto podría explicarse a un mayor empeño en las tareas defensivas como coberturas, carreras de recuperación e intercepciones (Bradley & Ade, 2018).

El fuera de juego resultó ser una variable que explica cuando el extremo gana los partidos y se puede interpretar como un indicativo de que el equipo estuvo permanentemente atacando al rival, aumentando así la probabilidad de tener muchas acciones reglamentarias de este tipo. Existe una clara relación positiva dentro del equipo con la probabilidad de ganar, sugiriendo que estos jugadores ganadores son más agresivos en el proceso de atacar (Liu et al., 2016; Zhou et al., 2018).

6.1.6. Centro delanteros

Los disparos, la finalización, la presión y el regate, son variables fundamentales que un centro delantero debe poseer (Modric et al., 2020). De acuerdo con los hallazgos encontrados en el presente estudio, algunas de estas variables fueron significativas para el modelo diseñado, como por ejemplo, los tiros a portería, por fuera o por dentro, y la finalización. Estos jugadores se caracterizaron por ser anotadores, asistentes y poseer habilidad para pasar el balón, que fueron todas variables encontradas y que según el modelo explican muy bien el óptimo rendimiento. Barron et al. (2020), destacan como fundamental la capacidad de anotar, ya que esta cualidad fue uno de los indicadores claves para explicar el 92.3% del modelo predictivo diseñado para clasificar la posibilidad de que el delantero compita al siguiente año en una liga más calificada.

Por otra parte, hubo dos variables técnicas se asociaron inversamente con el índice Instat[®], la efectividad de los centros (igual que el modelo de delantero extremo) y el regate. En cuanto al regate, podría considerarse lógico por estar ubicado en una zona del campo con una gran densidad de defensas rivales; sin embargo, la capacidad de conducir y regatear con el balón del delantero ha sido mencionado como un indicador clave de rendimiento (Hughes

et al., 2012). De la misma manera, la efectividad de los centros, fue un predictor significativo en el fútbol de Inglaterra para distinguir entre delanteros con diferentes niveles competitivos (Barron et al., 2020), ya que es común que los delanteros aprovechan sus movimientos para ubicarse en la espalda de los centrales o laterales y puedan ejecutar esta acción. Los resultados de este estudio concuerdan con el trabajo de Taylor et al. (2008), quienes aducen un aumento de ambas acciones cuando se pierde.

A nivel defensivo, la variable “disputas” permitió mejorar la puntuación, dejando claro que la lucha y la agresividad deben estar siempre presentes. En los últimos años, esta cualidad ha tenido una importante evolución en las posiciones cada vez más ofensivas (e.g., mediocampistas ofensivos y extremos delanteros) como producto del patrón defensivo característico en el estilo de juego moderno (Konefał, Chmura, et al., 2019a). La intención es recuperar el balón en el último tercio de la cancha, inmediatamente después de perderlo, con el objetivo de estar más cerca del marco rival y aumentar las posibilidades de anotar (Almeida et al., 2014; Vogelbein et al., 2014).

En cuanto a las variables físicas, dos características significativas fueron halladas. Primero y de forma positiva, los sprints absolutos que se atribuyen al continuo movimiento que estos jugadores ejercían. Estar situados cerca de la portería contraria significa que, a través de los sprints, pueden aparecer muchas oportunidades de realizar acciones de ataque (Modric et al., 2019). De hecho, estos jugadores presentaron según el modelo, pocas acciones defensivas en comparación con otras posiciones, lo que sugiere que realizaron más acciones de ataque que aumentaron la probabilidad de obtener resultados positivos en el juego. Segundo, la distancia recorrida a velocidades entre 12-19 km/h consideradas carreras submáximas que perjudicaron la nota final demostrando que en esta posición será clave el desarrollo de actividades de muy alta velocidad (Modric et al., 2019).

Para concluir esta primera parte del análisis de variables que explican el rendimiento de los futbolistas a partir del índice de rendimiento índice Instat[®], se logró calificarlos según su posición dentro del terreno de juego, identificando las variables más significativas, a través de la construcción de seis modelos de regresión lineal con una capacidad predictiva bastante alta. Sin embargo, no todo fue bueno, ya que se concluye que las variables físicas recogidas durante el partido por medio de los dispositivos GPS no fueron tan importantes. Al construir estos modelos, muchas variables no fueron consideradas estadísticamente significativas y las

que sí formaron parte tuvieron coeficientes bajos; es decir, aportaron muy poco a la obtención de la calificación final.

Esto se considera negativo porque se esperaba que estas variables lograran explicar gran parte del rendimiento de los futbolistas en el terreno de juego. Sin embargo, estos resultados coinciden en parte con los encontrados en el estudio de Modric et al. (2019), quienes no encuentran relación entre la condición física de los jugadores y el índice de rendimiento índice Instat[®]. A la vez, concluyen que los jugadores pueden ser muy eficientes en el desempeño de sus funciones técnico-tácticas en sus posiciones de juego, y por ende obtener valores elevados en dicho índice, a pesar de tener una condición física algo inferior. También mencionan que este índice de rendimiento podría verse influido por la experiencia del juego, ya que los jugadores de mayor edad son capaces de manejar sus exigencias futbolísticas de forma más eficiente.

Del mismo modo, el índice Instat[®], al ser un cálculo construido sobre la base de un conjunto único de parámetros técnico-tácticos clave para cada posición (entre 12-14 variables de rendimiento) (Modric et al., 2020) se podrá inferir que los resultados del presente estudio se verán influidos por la eficacia de esos elementos.

6.2. *Identificación de variables en la predicción del resultado del partido*

Con el objetivo de identificar las variables que influyeron en el resultado de un juego de acuerdo a la posición en el campo, se utilizó el aprendizaje automático o machine learning con modelos de aprendizaje supervisado de regresión y clasificación. Se utilizaron tres técnicas diferentes que consiguieron alcanzar el objetivo: a) la regresión logística, b) árbol de decisión, y, c) bosques aleatorios. En sí, los modelos ajustados, presentaron una baja-moderada capacidad predictiva; sin embargo, al ser un análisis exploratorio era algo que se podía esperar.

6.2.1. *Regresión logística*

Los seis modelos creados para cada posición expusieron una moderada capacidad predictiva de rendimiento. De esta manera, se lograron valores de 69.2% para defensas centrales, 68.3% para defensas laterales, 73.9% para mediocampistas defensivos, 69.6% para mediocampistas creativos, 72.2% para centro delanteros, y 75.1% para extremos delanteros, siendo éste el modelo con mayor capacidad predictiva. Además, es importante destacar que

todos los modelos cumplieron con los supuestos de normalidad, homocedasticidad, linealidad y no multicolinealidad, lo cual reduce la posibilidad de sesgos en la información analizada.

6.2.1.1. Defensas centrales

Entre las variables asociadas a la probabilidad de ganar según los análisis obtenidos en el presente estudio se destacan la cantidad de tiros y el porcentaje de efectividad en los lanzamientos. La cantidad de tiros, al igual que en el análisis anterior, fue un indicador importante de rendimiento para esta posición, ya que en situaciones de balón parado a favor (e.g., saques de esquina o tiros libres), estos jugadores tuvieron mayores posibilidades de ejecutarlos.

Ahora bien, cuando el defensor central generó muchos tiros “directos” a portería (realizados en dirección al marco) o “goles esperados”, la probabilidad de ganar disminuyó. Posiblemente en situaciones con marcador en contra o empate, la constancia de estos jugadores por ir al frente con la intención de buscar el partido, afectó al verse reflejado negativamente dentro del modelo estadístico.

Del mismo modo, los pases efectivos disminuyeron la probabilidad de ganar. En el análisis de regresión lineal para los defensores centrales, esta cualidad explicó muy bien el rendimiento en el índice Instat®; sin embargo, en esta técnica se investiga la influencia que tiene la efectividad del pase en el resultado final del partido. Esto se podría interpretar como la forma en cómo se atacó al rival, posiblemente a través de un método poco eficiente, al no arriesgar el balón en zonas más ofensivas, tratando de jugar a la “segura” a una altura del campo con poca marca (e.g., mitad de campo propia). Según Lago-Peñas et al. (2017), el estilo de “mantener la posesión” implica un juego más lento con movimientos defensivos y un menor riesgo en los pases. Otra razón puede ser que la mayoría de los equipos del campeonato, hacían cerrojos defensivos, jugando detrás de la línea del balón, lo que obligaba al central a realizar mucho pase horizontal, perdiendo la profundidad. Por lo tanto, se requiere que los jugadores defensores, no sólo tengan una capacidad de realizar pases acertados, sino también dirigidos hacia adelante (Barron et al., 2020; Bradley et al., 2013) y en zonas del campo específicas (Barron et al., 2020; Cintia et al., 2015).

Sin embargo, cuando se buscó el pase largo, el alto porcentaje de efectividad resultó ser la variable que mejor explica ganar en esta posición. La oportunidad de conectar con los

jugadores ofensivos más distantes, ubicados en la zona alta de ataque, crea muchas situaciones para alcanzar el marco contrario (Adams et al., 2013).

El uso de regates efectivos por parte de los jugadores centrales correlacionó inversamente con el índice de rendimiento, y esto puede deberse a que mientras se pierde hay un aumento en el regate (Taylor et al., 2008). Además, es una mala aplicación táctica, pues, se corre el riesgo por la ubicación que se tiene en el campo (última línea del sistema) de una pérdida de balón muy cercana a la propia portería.

En cuanto a las variables físicas, un aumento en la carrera de alta velocidad o HSR se asoció a una menor probabilidad de ganar, y esto puede deberse a la aplicación de un mayor número de acciones defensivas, como coberturas, cierres o interceptaciones, carreras de recuperación, que han sido identificadas por Bradley y Ade (2018) y que aportan entre un 26%, 16% y 20%, respectivamente, de la distancia total en HSR.

Este aumento en la carrera a alta velocidad también puede atribuirse en fase ofensiva cuando el defensa central recurre mucho al pase largo o juego directo, como estrategia ofensiva, que lo obliga a recorrer distancia con el fin de “achicar” el equipo y quitar espacio. En esta situación, se corre el riesgo de estar más propensos a recibir contras por quedar más expuestos y generalmente sucede con frecuencia cuando se pierde (Andrzejewski et al., 2018).

Otras variables que aportaron a explicar el rendimiento en los defensores, aunque con coeficiente pequeños fueron la distancia total recorrida y la distancia recorrida entre minutos (m/min). Un estudio realizado durante el Mundial de FIFA Brasil 2014, logró demostrar que los defensores cuando cubren una mayor cantidad de metros por minuto, sobre todo en posesión de pelota, el rendimiento del equipo mejora con una predicción estimada del 6% (Ugalde Ramírez et al., 2018).

Para terminar, la tarjeta roja y las faltas fueron medidas disciplinarias que afectaron en el rendimiento de los defensores centrales. La tarjeta roja conlleva la expulsión de un jugador, lo que debilita el rendimiento del equipo sancionado en términos de anotación de goles y resultado del partido, mientras que la sanción de tarjeta amarilla puede comprometer el rendimiento defensivo de un jugador para evitar su expulsión (Liu et al., 2016).

6.2.1.2. Defensas laterales

Las ocasiones generadas y los goles, fueron variables significativas que ayudaron a explicar la probabilidad de ganar para los defensas laterales. El estilo de juego utilizado promovía que estos jugadores tuvieran la oportunidad de subir por el carril con el fin de hacer daño al rival.

Algunas acciones recurrentes que generaron ocasiones de gol y su conversión fueron, por ejemplo, conducir el balón de afuera hacia adentro del campo en dirección al área grande con el fin de acercarse al marco e intentar rematar, o finalizar una situación de juego que progresaba por el costado opuesto, la cual la mayoría de las veces no tienen referencia de marca. Otra forma para generar opciones de gol fue jugar en posición de segundo volante defensivo, que les permitió ganar los rechazos del rival y tener el marco de frente, aumentando así las posibilidades de anotar.

Por otra parte, las cualidades que restaron a la probabilidad de ganar fueron los balones perdidos, los fuera de juego, el porcentaje de acciones efectivas y las disputas en ataque ganadas. Las dos primeras se pueden deducir y de alguna manera manifiestan la dependencia ofensiva que tenía el equipo con estos jugadores. Lo que pareciera difícil de explicar es cómo el porcentaje de acciones efectivas y las disputas en ataque ganadas no clasificaron a favor del gane; sin embargo, puede deberse al empeño que el jugador asumió para mejorar su rendimiento mientras el equipo perdía o empataba.

Las cualidades físicas como distancia recorrida por minuto (i.e., m/min), distancia recorrida a un rango de velocidad entre 19 a 24 km/h y la aceleración lograron clasificar positivamente y corroboran la relevancia que tienen como variables predictoras en el resultado del partido (i.e., ganar) cuando estos jugadores presentan altos puntajes en ellas (Andrzejewski et al., 2018; Bradley & Ade, 2018; Bradley et al., 2013). De igual manera, la aceleración es una característica fundamental en los defensas laterales por el tipo de movimientos explosivos que realiza para marcar y su vocación por correr a alta intensidad sobre el “callejón”. Geurkink et al. (2021), en su estudio con futbolistas profesionales concluyen sobre el impacto que tiene la capacidad de acelerar en los segundos tiempos en el resultado final del partido.

En el caso de las medidas disciplinarias, la tarjeta roja fue la principal variable que explicó la posibilidad de ganar, contrario a los defensas centrales. La lógica diría que jugar

con el equipo completo (i.e., 11 jugadores en el campo) aumentaría la probabilidad de ganar, por tener cubierto gran parte del campo; sin embargo, algunos estudios no han logrado asociar la tarjeta roja con ganar, empatar o perder (Kubayi & Toriola, 2020). Incluso se han relacionado las tarjetas rojas con el nivel de oposición y los resultados tienden a indicar que los equipos de nivel superior (e.g., el presente equipo estudiado) se ven menos influidos por la sanción de las tarjetas (Liu et al., 2016).

Al igual que con el análisis de los defensas centrales, la tarjeta amarilla aporta de manera negativa, limitando el accionar del jugador al tener mayor precaución a la hora de marcar y ser agresivo, aspectos que para un defensa lateral son importantes por las características propias de su perfil defensivo.

6.2.1.3. Mediocampistas defensivos

Los mediocampistas defensivos obtuvieron muchas más variables ofensivas significativas que defensivas para explicar cuando se ganó y entre ellas se encuentran los goles, goles esperados, asistencias, ocasiones generadas y porcentaje de efectividad de lanzamiento; demostrando un importante aporte en las acciones de ataque del equipo, a pesar de tener un rol más defensivo. La influencia en este tipo de acciones, se debe a su participación en la táctica fija (descrito en el análisis previo) y a los despliegues ofensivos en situaciones de partido como consecuencia de un intercambio en las funciones con el mediocampista ofensivo.

Sin embargo, la proyección del jugador al ataque no siempre calificó positivo, ya que algunas acciones restaron a la probabilidad de ganar; por ejemplo, al aumentar la efectividad de los regates y los tiros a portería. Estos resultados sorprenden ya que, a mayor cantidad de disparos a portería, mayor oportunidad de marcar goles (Kubayi & Toriola, 2020; Lago-Peñas et al., 2010; Lago-Peñas & Dellal, 2010) y por ende ganar los partidos. Además, en el análisis anterior que contempló el rendimiento individual de acuerdo al Índice Instat®, el aumento en ambos indicadores generó una mejor nota para esta posición. Por lo tanto, será necesario estudiar el momento del partido para comprender mejor estos resultados, ya que se ha logrado identificar una mejora significativa en las acciones técnicas como regates y remates a portería, cuando el equipo pierde (Taylor et al., 2008).

Las variables defensivas de interceptaciones y al porcentaje de disputas ganadas se asocian a la probabilidad de tener éxito durante el partido, y representan muy bien el rol que

el mediocampista defensivo cumple en el juego, esto por cuanto, se encarga de realizar labores de contención, de recuperación, de bloqueo, interrupción y anticipación de pases; y de destrucción del juego ofensivo del rival (Anaya Hernández, 2016). A pesar de esto, un alto porcentaje de robos de balón con éxito y un alto número de disputas en ataque ganadas disminuyeron la posibilidad de ganar y esto puede deberse a la gran cantidad de rebotes ganados o segundas jugadas ganadas en campo contrario, cuando el equipo está volcado al frente en situaciones de adversidad en el resultado (empate o pérdida). Casal et al. (2016) encuentran que la mayoría (48.9%) de las transiciones defensivas analizadas durante un estudio en el Mundial de Sudáfrica 2010, se iniciaron en el centro ofensivo del campo, precisamente la ubicación descrita anteriormente por parte del mediocampista defensivo en fase ofensiva, y que la mayoría de las recuperaciones de balón se producen en la zona defensiva posterior y en el mediocampo ofensivo (Lago-Peñas et al., 2011).

Por otra parte, el HSR afecta al modelo negativamente; es decir, que si se corre mucho la probabilidad de ganar disminuye. Esto podría ser consecuencia de un aumento de actividad en las funciones defensivas que los obligan a realizar acciones como carreras de recuperación, coberturas, interceptaciones, o carreras de aproximación para presionar al adversario, y han sido descritas anteriormente por Bradley y Ade (2018). También, puede deberse según el estudio de Chmura et al. (2018), a que los mediocampistas mientras ganan arriesgan menos al no verse obligados a realizar acciones ofensivas continuamente. Estos autores concluyen que los mediocampistas cuando ganan recorren una distancia significativamente menor a cuando pierden, en intensidades de carrera entre 17 y 24 km/h.

6.2.1.4. Mediocampistas ofensivos

Los volantes de creación son responsables de armar jugadas de ataque creando situaciones de gol con pases algunas veces incisivos y penetrantes, para que definan los delanteros y otras ocasiones mesurados e inteligentes que le rompan el ritmo al rival. También tienen la libertad de agregarse con frecuencia a la línea delantera, atacando por el centro o por el costado y participando activamente en la definición (Anaya Hernández, 2016). Por lo tanto, los goles, las asistencias y los tiros, son acciones técnicas muy comunes en el volante ofensivo y en el presente estudio lograron explicar cuando el equipo gana.

Según Castellano Paulis et al. (2012) las diferencias entre equipos ganadores y perdedores en el fútbol profesional se aprecian principalmente en el número y la efectividad

de los disparos a puerta, y los vincula a una de las tareas asignadas a los centrocampistas ofensivos. Como se ha descrito anteriormente, la construcción del juego pasa por este jugador, de ahí que cuando pierde mucho balón, resta también la probabilidad de ganar.

Desde el punto de vista defensivo, las disputas defensivas ganadas tuvieron un peso significativo en el modelo estadístico y muestran una disposición de estos jugadores para realizar labores de marca; sin embargo, un alto porcentaje de robos de balón con éxito, disminuyó la posibilidad de ganar. Estadísticamente, este resultado no concuerda con lo encontrado por Casal et al. (2016), quienes afirman que a mayor número de recuperaciones del balón, mayor posibilidad de ganar. Algunas posibles razones que explican esta discrepancia, sea el nivel de los equipos, el sistema de juego utilizado, la posición final en el torneo, así como una serie de variables que se escapan al análisis del presente estudio.

En cuanto a las carreras en diferentes intensidades, se logró identificar que un aumento en la distancia recorrida entre 12-19 m perjudica el resultado del partido. Con un rango de velocidad más delimitado (14-17 km/h), los resultados de Konefał, Chmura, Kowalczyk, et al. (2019) coinciden con los encontrados en el presente estudio, pues según esos autores, en su estudio en la Bundesliga alemana durante la temporada 2014/2015, cuando los mediocampistas incrementaban en 0.1 km la distancia recorrida, entre 14-17 km/h (promedio por partido= 1.03 ± 0.20 km), disminuían las posibilidades de ganar en un 16.6%.

En otro estudio de Chmura et al. (2018), no se encuentra una relación entre la distancia recorrida entre las velocidades 11-13.99 km/h, ni 14-16.99 km/h con el resultado final del partido para esta posición; sin embargo, logran demostrar que cuando los mediocampistas recorren menos distancia, entre 17-20.99 km/h, las probabilidades de ganar aumentan.

Por otra parte, cuando se ganó, los mediocampistas ofensivos alcanzaron grandes distancias en carreras a alta intensidad (19-24 km/h) y máxima velocidad que reflejaron ser determinantes para el modelo de juego desarrollado. De esta manera, para lograr comprender los resultados obtenidos en el presente análisis, se puede concluir que para aumentar las posibilidades de ganar se requiere incrementar los esfuerzos de máxima intensidad en contraposición con las carreras submáximas.

6.2.1.5. Extremos delanteros

Para este grupo de jugadores se obtuvo el modelo con el mayor rendimiento; sin embargo, algunas variables significativas que pueden ser consideradas vitales para los

extremos, más bien le restaron probabilidad de ganar. Por ejemplo, el aporte en cuanto a los goles, los tiros, los pases efectivos y los regates con éxito. Estas variables se han representado en el estudio de Taylor et al. (2008) como fundamentales en relación con el resultado final del partido.

Por su ubicación en la cancha, este jugador busca constantemente variabilidad en la forma de atacar para aumentar la dificultad de la oposición, y el regate es un arma importante para la impredecibilidad. Sin embargo, buscar el regate como un único método de ataque, puede llevar al juego hacia lo individual, privando consecuentemente la colectividad.

En el presente estudio las variables que aportan para ganar fueron las asistencias, los goles esperados, los tiros por fuera, las ocasiones generadas, las faltas recibidas, el porcentaje de acciones efectivas y el porcentaje de lanzamiento. Este último podría verse sumado por la efectividad que haya tenido este jugador en el pase largo, sobre todo buscando hacia el centro ya que por su ubicación (i.e., primera línea ofensiva) era poco probable contar con jugadores más adelantados que el extremo, en el campo de juego.

En cuanto a las variables físicas, aunque se encontraron efectos pequeños pero significativos para el modelo, un aumento en las carreras a alta velocidad, la distancia recorrida superior a 19-24 km/h y la velocidad máxima, se asociaron con la probabilidad de ganar. Según Konefał, Chmura, Kowalczyk, et al. (2019) en su estudio, los jugadores extremos alemanes recorrieron en promedio 0.42 ± 0.15 m a una velocidad de 24 km/h por partido. Cuando incrementaban en 0.1 km la distancia, las probabilidades de ganar aumentaban en un 31.7%.

Algunos estudios han caracterizado estos esfuerzos por ser muy largos, con mucha duración (Ade et al., 2016) y son considerados un requisito previo para un rendimiento exitoso en el fútbol, en acciones como marcar un gol (Faude et al., 2012), crear un disparo a puerta o eludir a un oponente (Oliva-Lozano et al., 2023).

Según Bradley y Ade (2018), los extremos cubren más distancia de HSR en el último tercio de la cancha e identifican cuatro acciones ofensivas en presencia del balón que propician el aumento en esta variable: a) driblar hacia el centro del campo, b) atacar la línea para vencer el fuera de juego del oponente, c) penetrar dentro del área rival, o d) avanzar corriendo por el carril. Estas acciones aportan un 32%, 12%, 10%, y 11%, respectivamente, de la distancia total de la carrera que realiza un jugador a alta intensidad. Mencionan que en

algunas acciones defensivas también recorren mayor cantidad de distancia a alta intensidad en comparación a otras posiciones (e.g., defensas, mediocampistas); por ejemplo, durante las interceptaciones (23% total de HSR). Este resultado pone una vez más de manifiesto la importancia del aporte defensivo que los extremo delanteros brindan durante la competencia en el fútbol moderno.

6.2.1.6. Centro delanteros

Con un 72.2% de precisión en el modelo, los goles, las jugadas de gol, los regates efectivos y un alto porcentaje de lanzamiento fueron acciones que se asociaron con la probabilidad de ganar. Las tres primeras se podrían deducir casi por lógica, en cuanto a la proximidad que tiene este jugador con el marco rival, ya que la mayoría de los esfuerzos terminan en el último tercio de ataque (Ade et al., 2016). En esta zona del campo, el delantero experimenta un continuo marcaje que lo obliga muchas veces a utilizar el regate como una herramienta fundamental, algunas veces para poder sostener el balón y esperar el acompañamiento de sus compañeros o cuando se arriesga a ir contra la portería rival.

El porcentaje de efectividad en el lanzamiento es un particular hallazgo, similar al encontrado con los extremos delanteros, ya que los atacantes juegan en zonas más avanzadas y tienen menos oportunidad de realizar pases hacia adelante (Barron et al., 2020). Sin embargo, puede deberse a una repetida permuta con el mediocampista, ampliando aún más la gama de recursos que debe tener el centro delantero y evidenciando una movilidad constante en el espacio ofensivo del equipo.

También las interceptaciones fueron vitales para ganar. La capacidad de robar balones en esta línea le permite tener una mejor oportunidad para acercarse al marco rival. Según Casal et al. (2016), cuando se pierde la posesión del balón, los equipos adoptan una estrategia defensiva persistente en el 57.7% de los casos, lo que indica la preocupación por recuperar rápidamente la posesión del balón y es, en estas situaciones, donde el delantero centro toma un papel fundamental. Los delanteros centro realizan carreras de alta intensidad, en arco, para canalizar a un adversario con el balón en una dirección mientras lo cierran con el fin de retrasar su ataque y permitir a sus compañeros de equipo apoyar la presión (Bradley & Ade, 2018).

En cuanto a las acciones que le restan posibilidades de ganar, realizar muchos pases efectivos no es una buena fórmula de ataque para esta posición, y puede explicarse como un

comportamiento predecible y poco arriesgado. El estilo de posesión de balón utilizado por el equipo analizado, pudo afectar el desenvolvimiento de este jugador, ya que la cantidad de pases y el número de pases acertados no necesariamente se encuentran entre los mejores predictores (Geurkink et al., 2021). Los autores encontraron en el ataque directo o el contraataque la forma más oportuna para poder ganar los partidos.

Los tiros fueron acciones que también afectan la probabilidad de ganar para esta posición. Este es un hallazgo inesperado, por cuanto es un jugador que siempre tiene presencia en el área rival y los disparos a puerta dentro al área de penal ha mostrado ser el mejor predictor de victorias según un estudio con jugadores profesionales belgas (Geurkink et al., 2021). De igual manera, Konefał et al. (2020) en el fútbol alemán, encuentran que las probabilidades de ganar el encuentro aumentan 11%, cuando un delantero realiza tiros con mayor frecuencia durante un partido. Una vez más, este análisis invita a reflexionar sobre la influencia de las variables situacionales durante el juego, especialmente al momento del partido, porque se han visto incrementos en los remates mientras se está perdiendo.

Otras variables que restaron al modelo obtenido fueron las faltas cometidas o las que recibieron. Esta última es interesante porque, al recibir muchas faltas, las posibilidades de ganar disminuyen y podría interpretarse como una acción defensiva del equipo rival para entorpecer o terminar un ataque sorpresivo.

Se debe destacar también, que cuando el delantero logró realizar distancias a intensidades de carrera superiores a 21 km/h, aumenta la posibilidad de ganar. Estas intensidades > 21 km/h suelen decidir sobre el resultado del partido (Bradley & Ade, 2018; Konefał, Chmura, et al., 2019a).

6.2.2. Modelos de Aprendizaje Supervisado o Machine Learning

Luego de realizar dos técnicas multivariadas (i.e., árboles de decisión y bosques aleatorios), se calcularon distintos indicadores de desempeño para cada modelo, precisión y AUC, con el fin de determinar en ambos casos cuál fue el que mejor se ajustó a los datos.

El estadístico de precisión es una métrica que se utiliza para determinar de todos los partidos que fueron clasificados positivos (“ganar”) qué proporción realmente lo fue; o sea, qué porcentaje de partidos están correctamente clasificados cuando se gana. El indicador puede estar entre 0.0 y 1.0, y cuanto mayor sea la puntuación mayor es el rendimiento del clasificador.

Por su parte, el AUC es la probabilidad de que el modelo clasifique un ejemplo positivo aleatorio más alto que un ejemplo negativo aleatorio, cuanto mayor es la puntuación mejor es el rendimiento del clasificador. Por ejemplo, si el AUC alcanza un valor aproximado de 0.5 (i.e., 50% de predicción), quiere decir que el modelo no tiene capacidad de discriminación para distinguir cuando el equipo gana o pierde. Esto sucedió con la técnica de árboles de decisión, en donde cuatro de los seis modelos no lograron alcanzar el 50%.

Si bien es cierto, las estimaciones fueron bajas-moderadas, en este tipo de clasificación el azar desempeña un papel importante en los goles y por ende en el resultado final del partido. Cuando la diferencia final de goles entre dos equipos es pequeña y el gol tiene un impacto en el resultado del partido, las precisiones cercanas a 1.0 parecen improbables (Geurkink et al., 2021).

En el presente estudio, el método que obtuvo mayor precisión en la estimación de la clasificación del resultado de partidos de fútbol (ganar/no ganar) fue el modelo de bosques aleatorios, por lo que se discutirán los principales indicadores obtenidos según la posición del jugador dentro del terreno de juego.

6.2.2.1. Defensas centrales

Las variables que tuvieron la mayor importancia fueron el número de sprints, el número de pases efectivos, el HSR, el promedio de acciones efectivas, los goles esperados y el porcentaje de disputas defensivas ganadas.

Las variables técnicas más significativas para la base de los defensas no fueron propiamente las acciones defensivas. Los pases efectivos y el promedio de acciones efectivas, ponen de relieve su importancia cuando se ganan los partidos. Para Adams et al. (2013), los equipos exitosos tienen defensores con recursos técnicos superiores que les permite desplazar a la oposición por el campo con pases efectivos como parte de la estrategia, esperando a que surjan oportunidades de ataque.

En el estudio de Adams et al. (2013), el indicador pase efectivo fue catalogado como el factor que más diferenció entre los mejores equipos posicionados en la tabla final (4 primeros) en relación a los últimos (4 últimos) en la Liga Premier para los jugadores ubicados en esta posición. De acuerdo con su análisis, los equipos peor ubicados se caracterizaron por defensas que jugaban “directo” (e.g., pase largo) como una solución a la falta de calidad técnica de los jugadores al verse presionados por sus rivales. Un similar hallazgo fue

reportado por Ugalde Ramírez et al. (2018), en el que se reporta que los defensas de los ocho mejores equipos del Mundial Brasil 2014 completaron mayor cantidad de pases efectivos que los defensas de los ocho equipos ubicados en las últimas posiciones.

Los goles esperados es otra variable ofensiva importante que explica el resultado para los defensas centrales y ha sido identificada como indicador predictivo del rendimiento (Barron et al., 2020; Geurkink et al., 2021; Ugalde Ramírez et al., 2018), que se manifiesta, sobre todo, en situaciones de táctica fija y finalizadas con la cabeza (Ade et al., 2016). Los goles esperados toman en cuenta la ubicación de los disparos y calculan qué tan probable es que el disparo sea convertido en gol.

El porcentaje de disputas defensivas ganadas representa una actividad clave para los jugadores en defensa (Konefał, Chmura, Kowalczyk, et al., 2019; Konefał, Chmura, et al., 2019a, 2019b), y esta idea se ha visto confirmada por la expansión constante y evolutiva de dichas actividades en temporadas sucesivas de la Bundesliga alemana (Link & de Lorenzo, 2016).

En cuanto a las variables físicas, el número de sprints y la distancia a alta intensidad, fueron hallazgos llamativos en esta posición, ya que según la literatura consultada, los defensores centrales no sólo promedian los menores puntajes de carreras a alta intensidad en comparación con el resto de posiciones (Andrzejewski et al., 2018; Bradley & Ade, 2018; Chmura et al., 2018; Oliva-Lozano et al., 2020), sino que, cuando ganan realizan un menor número y distancia de sprint en fase defensiva (Andrzejewski et al., 2018). Esto se comprende por la necesidad de mantener al equipo compacto, con poca distancia entre los jugadores, limitando el espacio a la oposición.

Los resultados obtenidos en el presente estudio podrían ser explicados por el estilo de juego utilizado por el equipo, que daba prioridad al juego ofensivo con el fin de ganar los partidos. Las líneas se adelantaban y se forzaba a los defensores a ubicarse en la mitad del campo, con la finalidad de jugar la mayor parte del tiempo en zona rival. En esta posición, el papel fundamental de los defensores era estar permanentemente en vigilancia a los posibles receptores adversarios, reconociendo que a sus espaldas los rivales contaban con un gran espacio. Esta situación particular de juego, promovía carreras de alta intensidad cuando el equipo era sorprendido por contraataques del rival que obligaban al repliegue rápido.

Otra manera de sumar distancia a alta intensidad ocurría; por ejemplo, cuando el defensor lateral era sobrepasado por el rival con balón, que generaba una cobertura inmediata del defensa central más próximo como aplicación táctica a los movimientos defensivos en la línea (i.e., dos defensores centrales más dos defensores laterales). De acuerdo con Bradley y Ade (2018), tanto las carreras de recuperación como las coberturas representan una demanda física plasmada en las distancias alcanzadas con carreras a alta intensidad, entre un 20% y un 31%, respectivamente.

También se debe considerar que una pequeña parte del total de las carreras a alta intensidad (5-10%) están cubiertas cuando el balón no está en juego (Bradley & Ade, 2018), como respuesta a la reubicación dentro del sistema inmediatamente después de una situación de partido o cuando el defensor recorre todo el campo con velocidad en situaciones de pelota quieta o balón parado a favor (e.g., desplazamientos previos a un saque de esquina), como frecuentemente ocurría durante los partidos observados.

Las aceleraciones y las distancias recorridas a velocidades entre 12-19 km fueron variables importantes que se asociaron al éxito. Ambos indicativos concuerdan con los resultados hallados con jugadores profesionales en Croacia, que encontraron que para la posición de defensa central tanto la distancia de carrera entre 14.4-19.7 km/h como la aceleración a alta intensidad, tuvieron gran impacto en el marcador final de los partidos (Modric et al., 2019). Este estudio manifiesta una importancia mayor a la aceleración, por estar ligada a la capacidad de ganar duelos y presionar a los jugadores contrarios; una tarea muy característica de los defensores centrales y que puede ser entrenada y mejorada por medio de la preparación física.

6.2.2.2. Defensas laterales

Entre las variables más importantes identificadas con el modelo de bosques aleatorios, se encontraron la velocidad máxima, las faltas recibidas, el porcentaje de disputas ganadas, el número de sprints, el porcentaje de efectividad de los lanzamientos, la distancia cubierta entre 19-24 km/h, y las ocasiones de gol generadas.

Como se puede apreciar, las variables físicas de carrera a alta intensidad (e.g., velocidad máxima, número de sprint, distancia entre 19-24 km/h) tuvieron una gran contribución para explicar cuando se gana.

Un defensa lateral recorre cerca de 600 m por partido a intensidades ≥ 21 km/h (Bradley & Ade, 2018), y alcanza una mayor distancia total de sprint que el resto de las posiciones (Andrzejewski et al., 2018; Konefał, Chmura, Kowalczyk, et al., 2019; Konefał, Chmura, et al., 2019a). Andrzejewski et al. (2018) lograron demostrar que la distancia de sprint es mayor cuando se gana que cuando se pierde, concordando con los resultados del presente estudio.

Estas acciones generalmente se asocian más en ataque, por la insistencia en situaciones como desdobles o ataque por las bandas (Andrzejewski et al., 2018; Bradley & Ade, 2018) que finalizan en el último tercio del campo (Ade et al., 2016).

El perfil ofensivo de los laterales y su capacidad para desarrollar desplazamientos muy intensos le permitía generar acciones ofensivas como goles esperados, ocasiones generadas y asistencias, que tuvieron un significativo aporte al modelo estadístico y que generalmente son ejecutadas con esfuerzos previos a la máxima intensidad.

Otra variable significativa fue el porcentaje alto de regates con éxito, al igual que en el modelo de regresión lineal para explicar el rendimiento del lateral en relación al índice de rendimiento, y pone en evidencia la necesidad de contar con laterales que tengan un buen desempeño en el 1 vs. 1 (Ade et al., 2016; Modric et al., 2019).

En otras palabras, para aumentar las probabilidades de ganar un juego, el equipo requirió en fase ofensiva la participación constante del defensa lateral. De hecho, las faltas recibidas tuvieron una gran importancia para explicar cuando el equipo gana.

6.2.2.3. Mediocampistas defensivos

El modelo de los mediocampistas defensivos tuvo una capacidad de predicción bastante baja, del 47%; sin embargo, se hallaron variables importantes que han sido muy correlacionadas con el éxito en esta posición.

El número desaceleraciones (-10 a -3 m/s²) fue el indicador con mayor importancia para el modelo que explica cuando el mediocampista ganó sus partidos. Actualmente, los jugadores realizan más acciones cortas de alta intensidad para cumplir con las responsabilidades tácticas (Bush et al., 2015), mostrando una progresión evolutiva en la frecuencia de las aceleraciones rápidas (Bush et al., 2015; Geurkink et al., 2021) pero, sobre todo, una prevalencia de las desaceleraciones de alta intensidad para el rendimiento en los partidos de fútbol, en los que se han encontrado distancias recorridas de aproximadamente

162 m, muy superiores a otros deportes como el rugby y el fútbol australiano (54 m y 149 m, respectivamente) (Harper et al., 2019).

Las desaceleraciones imponen una demanda significativa en la capacidad de absorber repetidamente altas fuerzas excéntricas de frenado (Springham et al., 2020) que pueden ser relacionadas en estos jugadores por su rol en la marca y el tener que estar cambiando constantemente de dirección por su ubicación centralizada en la cancha (Konefał, Chmura, et al., 2019a).

Las variables distancia recorrida absoluta (m) y distancia recorrida relativa al tiempo (m/min), resultaron ser métricas importantes para explicar el modelo en esta posición. Se sabe que los mediocampistas recorren más distancia total que los defensores y delanteros (Vigne et al., 2010), además de que lo hacen a gran intensidad (Bradley & Ade, 2018). También se ha encontrado que el sistema de juego utilizado por el equipo del presente estudio (i.e., 1-4-3-3), demanda mayor distancia de carrera para esta posición, aproximadamente 11% más, en comparación con otros sistemas, como por ejemplo, el 1-4-4-2 (Tierney et al., 2016), con en el que con dos mediocampistas defensivos se pueden distribuir el espacio.

Cuando están en fase ofensiva, recorren la mayor cantidad de metros debido a que se involucran constantemente en desplazamientos para desmarcarse y apoyar a los demás compañeros y contribuir en la elaboración del juego, por medio, de pases cortos o medios (Anaya Hernández, 2016). Según Ugalde Ramírez et al. (2018), una larga distancia recorrida por unidad de tiempo (i.e., m/min) aumenta su contribución al éxito del equipo.

El porcentaje de acciones efectivas también mejoran su puntuación, y concuerda con lo descrito por Konefał, Chmura, et al. (2019a) en su análisis de las actividades técnicas distinguiendo entre centrocampistas. Estos autores revelaron que los centrocampistas defensivos completaron un mayor número de pases con gran precisión en comparación con los centrocampistas ofensivos. Esta capacidad de generar muchos pases efectivos conduce a un mayor número de ocasiones de gol (Bradley & Ade, 2018; Konefał, Chmura, et al., 2019a).

Llama la atención cómo la velocidad máxima y los sprints explicaron muy bien el modelo, ya que son jugadores que, al estar centralizados, tienen menor espacio para desarrollar estas velocidades en comparación con los que juegan por afuera (Ingebrigtsen et al., 2015). Para diversos autores (Konefał, Chmura, Kowalczyk, et al., 2019; Konefał,

Chmura, et al., 2019a), los centrocampistas defensivos, en comparación con los volantes ofensivos, cubren distancias más cortas a muy alta intensidad, y realizan también un menor número de sprints, y menor velocidad máxima. Sin embargo, durante los partidos analizados del presente estudio fue común ver a estos jugadores haciendo coberturas y carreras de recuperación a una alta intensidad, trabajos muy propios de la línea defensiva (Bradley & Ade, 2018).

6.2.2.4. Mediocampistas ofensivos

Entre las variables relevantes que se pudieron identificar con el análisis de bosques aleatorios se encuentran los sprints, los pases efectivos, las acciones efectivas, el HSR distancia, la velocidad máxima y la distancia entre 12-19 km/h.

Al igual que con el modelo creado para los mediocampistas ofensivos en el análisis de regresión logística, las carreras de alta intensidad tuvieron un aporte significativo. Se ha mencionado que el aumento de las distancias recorridas a altas intensidades (i.e., HSR, sprints) y carreras a máxima velocidad por parte de estos futbolistas, se correlacionan con el resultado positivo del juego (Andrzejewski et al., 2018). Además, estas acciones se producen con mayor frecuencia en el juego ofensivo (Andrzejewski et al., 2018; Chmura et al., 2018), y cuando se logran sostener en el segundo tiempo, las posibilidades de ganar aumentan (Geurkink et al., 2021).

La carrera submáxima también explicó bien cuando se gana, en distancias recorridas entre 12-19 km/h. Se ha encontrado que un aumento de la distancia recorrida entre 14-17 km/h representa un cambio del 8.81% en las probabilidades de ganar, lo que se puede justificar en la necesidad del centrocampista a desplazarse sobre el terreno de juego con el fin de observar y percibir más cercanas las situaciones del partido, y para garantizar una ejecución más eficaz de sus acciones (Konefał et al., 2020).

Otras variables como las acciones efectivas y los pases efectivos, demuestran un importante peso ofensivo, ya que, al tener un mayor valor de centralidad de proximidad (i.e., lo cerca que está de otros compañeros), centralidad de grado (i.e., mayor número de pases), y centralidad de intermediación (i.e., mayores conexiones entre un jugador y otro) (Aquino et al., 2020), la eficacia en estas acciones, promueven a nivel colectivo una tenencia del balón más prolongada. Aunque las pruebas acerca de obtener una ventaja al dominar la posesión del balón manteniéndolo durante una parte más larga del partido en comparación con el rival

siguen sin estar claras (Adams et al., 2013), se ha logrado identificar que los pases y su efectividad, tanto en los defensas como en los mediocampistas tienen mucha influencia en el éxito del partido.

En cuanto a la aceleración, si bien es cierto en el estudio de Borghi et al. (2021) con jugadores élite sub-19, no se encontraron diferencias sustanciales en $Acc > 3 \text{ m/s}^2$ entre las posiciones en el sistema 4-3-3, mencionan que los volantes ofensivos son los que mayor número de aceleraciones realizan por partido.

6.2.2.5. Extremos delanteros

Las variables con mayor importancia para el modelo de los extremos delanteros fueron la velocidad máxima, la distancia, las entradas, los balones perdidos, y la carrera en m/min.

El indicador de rendimiento de máxima velocidad actuó como el de mayor importancia en el análisis del rendimiento para los jugadores que se desempeñan en esa posición. La arrancada explosiva, la capacidad de aumentar la velocidad a lo largo de la distancia recorrida y la capacidad de alcanzar rápidamente la velocidad máxima de carrera son componentes cruciales que afectan a la eficacia de los jugadores durante un partido de fútbol (Konefał, Chmura, Kowalczyk, et al., 2019; Konefał, Chmura, et al., 2019a).

Algunos investigadores (Oliva-Lozano et al., 2020; Oliva-Lozano et al., 2023) encuentran que los extremos realizan la mayoría de los sprints durante la ofensiva y en campo opuesto. Para el defensor rival, el hecho de tener que marcar a un jugador rápido le genera grandes complicaciones, al tener que decidir si dejar que reciba el balón o marcar de cerca sabiendo del gran espacio que pueda ser aprovechado. De ahí que un jugador capaz de superar a su adversario en una distancia de pocos metros es un gran activo para cualquier equipo de fútbol (Dellal et al., 2011). Andrzejewski et al. (2018) analizaron en la Bundesliga la temporada 2014/2015, la distancia de sprint recorrida por los extremos y encontraron distancias significativamente mayores en los partidos ganados que en los perdidos.

Otras cualidades que resultaron significativas para el desarrollo del modelo del presente estudio fueron la distancia recorrida y los m/min recorridos. En la actualidad, los extremos delanteros deben cubrir largas distancias, ya que en posesión del balón suelen participar en acciones ofensivas dinámicas en el último cuarto del campo y ejercer tareas defensivas para acompañar al defensor lateral. La distancia total de carrera ha sido asociada

al rendimiento de los jugadores (Chmura et al., 2018; Hassan et al., 2020; Konefał et al., 2020), especialmente la realizada con la posesión del balón (Hoppe et al., 2015). Sin embargo, para ganar, también se debe defender y se ha comprobado que estos jugadores durante el partido son los que recorren la mayor cantidad de carreras de recuperación que las otras posiciones (Ade et al., 2016).

Los resultados también arrojaron que variables con un enfoque defensivo destacaron entre las más importantes para explicar cuando se gana, apoyando el comentario anterior de la necesidad de que los extremos tengan esa doble función. Las entradas y las interceptaciones tuvieron un efecto positivo y fueron mencionadas en el análisis de regresión lineal para esta posición.

Por otra parte, cuanto menos balón pierda el extremo, aumenta la probabilidad de ganar el partido.

6.2.2.6. Centro delanteros

Entre las variables que se lograron identificar con mayor importancia y que pueden ayudar a predecir el resultado del partido se encuentran la velocidad máxima, el sprint, la distancia, la distancia entre el tiempo (m/min), la distancia entre 12-19 km/h, las acciones efectivas, la aceleración, la desaceleración y los pases efectivos.

Las probabilidades de ganar son mayores en los partidos en los que un delantero recorre un mayor porcentaje de distancia a velocidades >24 km/h y realiza un mayor número de sprints (Andrzejewski et al., 2018; Chmura et al., 2018). Esta cualidad ha evolucionado a lo largo de las temporadas para esta posición (Bradley et al., 2013) y se ha descubierto que cada aumento de 1% se asociará a un aumento constante de 4.08% en las probabilidades de ganar en las temporadas siguientes (Konefał et al., 2020).

En fase ofensiva, al igual que los extremos, estos jugadores requieren desplazamientos rápidos de alta intensidad, por lo que entre más veloces e intensos sean estos movimientos, más opciones y ventajas tendrán los delanteros sobre los defensas, aumentando las posibilidades de anotar (Faude et al., 2012; Konefał et al., 2020). Acciones como atacar un espacio para recibir un lanzamiento de pase, recibir centros de costado, o regatear rápido en el área de penal son patrones característicos en esta posición (Andrzejewski et al., 2018) y requieren alta capacidad para acelerar y frenar debido a los cambios bruscos de dirección (Harper et al., 2019). También en defensa, suman distancia en velocidad ya que les

corresponde iniciar las acciones para recuperar el balón perdido, como entradas o interceptaciones (Bradley & Ade, 2018).

Los metros por minuto recorridos fueron la segunda métrica de mayor importancia para el modelo estadístico del presente estudio. Dos situaciones pudieron haber inducido a este factor; por ejemplo, a nivel ofensivo, la necesidad de estar continuamente en movimiento ante defensas muy compactas o cerradas (ocurrido durante la mayoría de los encuentros), con el fin de crear y aprovechar los espacios otorgados por el rival. O también, defensivamente, al tener que presionar en campo rival con el propósito de robar el balón cerca del área rival. De hecho, Ugalde Ramírez et al. (2018) encontraron que en el mundial de Brasil 2014 el 19% del rendimiento de los equipos fue explicado por la gran cantidad de m/min alcanzados por estos jugadores, demostrando que el delantero moderno requiere una movilidad continua.

Se debe tomar en cuenta que esta movilidad tendrá que ser bien dosificada pues según Konefał, Chmura, Kowalczyk, et al. (2019), aumentar 0.1 km la distancia recorrida durante el primer tiempo, reduce 12.3% las probabilidades de ganar. En ese estudio los delanteros alemanes recorrieron cerca de 5.55 ± 0.35 km en la primera mitad de un partido.

La distancia recorrida entre 12 a 19 km/h también fue considerada como variable predictora, aduciendo que entre más distancia se recorra a esta velocidad (considerada carrera submáxima) también existe una mayor probabilidad de ganar los encuentros. De alguna manera, estos resultados podrían ser explicados de acuerdo a la posición de (Andrzejewski et al., 2018), en la que los delanteros centrales no asumen un papel protagónico en la marca y por lo tanto mantienen puntajes altos a estas intensidades de carrera.

Por otra parte, en cuanto a las acciones técnico-tácticas, el porcentaje de acciones efectivas (i.e., indicador de eficacia de todas las acciones que realiza el jugador durante el partido), que tuvo un peso importante en el modelo, puede reflejarse en la cantidad de indicadores significativos en el análisis; por ejemplo, los pases efectivos, los regates con éxito, las ocasiones de gol y los tiros por fuera. Cabe destacar que la variable gol, no tuvo un peso importante dentro del modelo para esta posición, un hecho que podría considerarse elemental, que se puede atribuir a la distribución dentro del campo, ya que sólo un atacante está centralizado en el sistema 1-4-3-3, aumentando la posibilidad de repartir los goles entre los demás jugadores del equipo (Adams et al., 2013; Barron et al., 2020). Esto refleja que los requisitos de la posición podrían estar cambiando.

Otro aspecto importante que no fue mencionado en ningún momento del presente trabajo fue que, de acuerdo a los resultados de los modelos de clasificación, se determinó que las variables importantes no se diferencian según el torneo, ni según el tipo de competencia (i.e., torneo nacional o internacional), ni la condición de localía.

En cuanto a la localía, debía esperarse un resultado diferente ya que la importancia de jugar en casa en el fútbol ha sido destacada en muchos estudios previos (Lago-Peñas et al., 2010; Lago-Peñas et al., 2011; Liu et al., 2016), y aunque esta ventaja ha disminuido constantemente desde un 70% en 1890 hasta alrededor del 60% en la actualidad, sigue estando claro que la ubicación del partido ejerce una gran influencia en el resultado final (Konefał et al., 2020). En el meta análisis de Jamieson (2010) se explica cuáles factores están asociados a esta condición y destacan, por ejemplo, el apoyo del público local, la fatiga del viaje del rival, la familiaridad en las condiciones locales, el sesgo del árbitro hacia el equipo local, la territorialidad, y los factores psicológicos. De igual manera, (Konefał et al., 2020) logran demostrar de forma más detallada cómo las probabilidades de ganar en casa son diferentes para cada posición de juego.

También se concluye en este segundo análisis, que las variables físicas de carrera de alta intensidad, como la velocidad máxima, el número de sprint (velocidades >24 km/h), el HSR (velocidades >21 km/h) resultaron vitales en todas las posiciones para poder ganar los partidos. El estilo de juego pudo haber contribuido a acumular distancias de alta velocidad, ya que, a mayor tiempo de posesión del balón, más intentos de encontrar espacios vacíos para crear muchas situaciones de gol (Asian Clemente et al., 2019).

En cuanto a la variable distancia total recorrida, apareció como indicador importante con un nivel de significancia bajo en el modelo de bosques aleatorios en cuatro de seis posiciones de juego (extremos y mediocampistas). Es común escuchar entrenadores y prensa destacando la importancia de esta cualidad para ganar los partidos; sin embargo, en el presente estudio la distancia recorrida a alta intensidad fue un indicador superior y más sensible del rendimiento.

En cuanto a los análisis estadísticos, se puede concluir que los resultados no fueron consistentes cuando se aplicaron las técnicas de regresión logística y bosques aleatorios. Para el análisis de regresión logística, las variables más significativas o que lograron explicar mejor la varianza, fueron las de naturaleza técnica, mientras que el modelo de bosques

aleatorios, hubo una predominancia de los indicadores físicos. Esto es un problema que debe ser resuelto por el análisis en sí de los datos, por lo que no es posible llegar a una conclusión definitiva.

Capítulo VII. Conclusión y Recomendaciones

1. Conclusiones con respecto a los objetivos específicos

Se elaboraron modelos de regresión lineal múltiple con el propósito de cuantificar el rendimiento de los jugadores según su posición de acuerdo a variables registradas durante los partidos y modelos de regresión logística, árboles de decisiones y bosques aleatorios para encontrar las variables que aumentan las probabilidades de ganar un partido. Como resultado de estos análisis, todos los modelos construidos resultaron ser capaces de asignar una calificación a los jugadores según su rendimiento en el campo y de identificar las variables que son influyentes en el resultado de un encuentro de fútbol, independientemente de la técnica utilizada; sin embargo, las mejores estimaciones se alcanzaron con bosques aleatorios, que es considerado un algoritmo de aprendizaje certero para el análisis de conjuntos de variables grandes. Además, permitió crear estimaciones de los indicadores más importantes para la clasificación, es decir, para predecir el resultado del partido.

De esta forma el enfoque específico por posición de este estudio permitió conocer un perfil del rendimiento físico-técnico individual de los futbolistas costarricenses, específicamente para cada posición de juego.

A continuación, se mencionan las variables más importantes que explican el rendimiento de un futbolista y las variables que aumentan las probabilidades de ganar un partido según la posición de juego (Tablas 16 y 17).

Tabla 16. Síntesis de las variables que explican el rendimiento del futbolista en el juego según posición.

Posición	Variables
Defensas centrales	Goles, Asistencias, Tiros, %disputas defensivas ganadas, entradas efectivas.
Defensas laterales	% efectividad pase, centros, disputas defensivas ganadas, % disputas por arriba ganadas, % de regates con éxito.
Mediocampistas defensivos	% efectividad de los centros, oportunidades de gol con éxito, asistencias, tiros, porcentaje de regates con éxito, entradas efectivas, interceptaciones.
Mediocampistas creativos y mixtos	% efectividad de pases, ocasiones de gol, % conversión, tiros a portería, asistencias esperadas, disputas por arriba ganadas, disputas defensivas ganadas.
Extremos	% disputas ganadas, goles, % efectividad de pases, asistencias esperadas, tiros a portería, centros efectivos, regates.
Delanteros centro	% efectividad de los centros, goles, asistencias, ocasiones de gol, % conversión, tiros, pases de finalización efectivos.

Tabla 17. Síntesis de las variables que aumentan las probabilidades de ganar un partido según la posición.

Posición	Variables
Defensas centrales	Número de sprints, número de pases efectivos, distancia recorrida a alta intensidad (HSRm), % de acciones efectivas, goles esperados
Defensas laterales	Máxima velocidad, faltas recibidas, % de disputas ganadas, número de sprints, % efectividad en lanzamiento
Mediocampistas defensivos	Número desaceleraciones (-10 a -3 m/s ²), m/min recorridos, % de acciones efectivas, máxima velocidad, interceptaciones.
Mediocampistas creativos y mixtos	Número de sprints, pases efectivos, % de acciones efectivas, distancia recorrida a alta intensidad (HSRm), máxima velocidad.
Extremos	Máxima velocidad, distancia recorrida (m), entradas, balones perdidos, m/min recorridos.
Delanteros centro	Máxima velocidad, m/min recorridos, % de acciones efectivas, distancia recorrida entre 12 a 19 km/h, número de sprints.

Se debe destacar que los resultados de este estudio no pueden distinguir completamente si una variable es la causa de un marcador favorable, o el efecto. Estos son análisis de regresión o de asociación, que explicaron lo que sucedió, pero no pueden predecir el futuro. De hecho, el diseño del estudio es retrospectivo, lo que indica que trata de explicar lo que sucedió durante los torneos analizados. Por otra parte, los análisis sirven para comprender las variables físicas y técnicas que intervienen en el fútbol, de manera que pueda aportarle una perspectiva nueva al staff técnico de un equipo para tomar decisiones y brindar retroalimentación del rendimiento a los involucrados en el juego.

2. Recomendaciones

Con el fin de explorar soluciones estadísticas a los datos que se tienen, la primera recomendación se centra en construir una base de datos que tenga una estructura estandarizada con el fin de que estén todas las variables que se quieren estudiar y que no haya faltante en los datos. Esto fue un primer problema que se presentó en los formatos de archivo, que consumió tiempo para poder corregir para tener una base de datos de calidad.

También dentro de esta categoría de recolección de la información, será vital, revisar que las posiciones de los jugadores sean correctamente asignadas, ya que es posible que sean etiquetados erróneamente en otra posición por el proveedor de la información (Instat[®]). Del mismo modo, habría que analizar la reubicación de los jugadores por el entrenador durante

el partido, como respuesta a lesiones, expulsiones o exigencias del sistema táctico. Los jugadores pueden experimentar diferentes cargas externas e internas cuando cambian de una posición a otra con diferencias de rendimiento.

En cuanto a los análisis estadísticos, es preciso tomar en consideración que los métodos de clasificación son robustos con una gran cantidad de casos, obteniendo predicciones o estimaciones muy certeras de lo que se quiere explicar; pero, cuando la cantidad de datos no es suficiente pueden presentar problemas de estimación, por lo que será necesario añadir nuevas observaciones a partir de técnicas de remuestreo (e.g., Bootstrap) (Moncada Jiménez et al., 2023), que permiten obtener resultados más robustos que respondan al objetivo del estudio.

Otro ejemplo, tiene que ver con la dimensionalidad por tener más variables que observaciones (multicolinealidad), como consecuencia de la gran cantidad de datos derivados a los que se debe depurar para evitar conclusiones erróneas. Se debe reducir dimensiones (e.g., análisis de componentes principales); es decir, tener cinco variables y reducirlas a una sola que explique o represente a todas, como por ejemplo, algunos indicadores como distancias a alta velocidad y porcentaje de distancia a alta velocidad, que entre ellas podría haber una alta correlación para entrar a una sola categoría.

En relación con la interpretación de los datos, si bien es cierto, los modelos muestran las variables de juego asociadas a la probabilidad de ganar, sólo se pudo dar seguimiento durante los partidos a un equipo, sin entrar en detalle en cuanto a la estrategia técnico-táctica empleada. En realidad, se estudiaron “resultados relacionados con el juego” relativamente simples (i.e., rendimientos de carrera obtenidos por GPS e indicadores de rendimiento de juego obtenidos con el InStat[®]), y no se incluyeron indicadores más complejos como la interacción, la cooperación y la oposición, manifestados en un partido de fútbol. Por ejemplo, los tiros de gol fue una variable significativa en muchos modelos, pero en sí, no informa sobre cómo crear oportunidades para marcar goles, y así desarrollar estrategias para mejorar los objetivos. Será necesario para futuras investigaciones evaluar los momentos específicos que están ocurriendo durante el juego, sea situación ofensiva o defensiva, considerando, por ejemplo, el uso de la imagen o video. De esta manera, se podrá hacer el análisis más preciso en la interpretación de los datos de acuerdo al tramo de video en el que se manifiestan.

Para finalizar y como consecuencia de la metodología que se presentó, es necesario comprender la importancia que tienen dos áreas profesionales para el manejo y la interpretación de toda esta información.

Desde el punto de vista de los profesionales en Estadística, se determinó que un analista de datos o científico del deporte (Gleason et al., 2023) que ingresa a una organización deportiva debe tener control total sobre los dispositivos de GPS Tracking, como los WiMu utilizados en el presente estudio, y mantener una constante comunicación con el cuerpo técnico para saber sus necesidades y preferencias. Además, esta persona se debe encargar de darle un sentido a la información recopilada por medio de reportes post partido, análisis descriptivo e inferencial sobre el rendimiento de los jugadores, contribuir en el análisis de vídeo, elaborar y manejar bases de datos y generar información sobre el próximo rival. Es importante recalcar que el analista de datos no se debería limitar solo a trabajar con el área deportiva del club, ya que sus conocimientos pueden ser útiles para el análisis y prevención de lesiones, el desarrollo de marketing, ajustar aspectos económicos y la ejecución de fichajes. El o la analista de datos debe tener un conocimiento futbolístico básico, habilidades de programación en software capaz de analizar y procesar datos, conocimientos estadísticos avanzados, predisposición y proactividad, habilidades comunicativas, relacionales y versatilidad. Los insumos necesarios para que la persona se desempeñe en esta posición son servidores locales y en línea, servicios de web “cloud” para poder analizar muchos datos al mismo tiempo, dispositivos de GPS Tracking capaces de medir con alta precisión las variables de interés y la disposición de una conexión a internet estable y potente.

Desde la perspectiva del profesional en Ciencias del Movimiento Humano debe planificar, analizar, supervisar y evaluar la calidad del entrenamiento deportivo con base a la información obtenida, por lo que no sólo requerirá de conocimientos en áreas como la anatomía, la fisiología, la biomecánica, la psicología, la sociología y la didáctica, sino también conocimientos en computación e informática y estadística. La comunicación de la ciencia que se realiza en el país también es importante, pues permite aportar al conocimiento internacional y recibir retroalimentación de otros colegas. El autor de esta tesis pudo comunicar dos de sus estudios realizados durante el periodo de estudios de posgrado (Anexo 3). La aplicación de estos conocimientos y la transferencia de la información a los deportistas, es una tarea crucial para ayudar a influir positivamente en la práctica. El científico deportivo

debe comprender y complementar las filosofías de los entrenadores, con la monitorización de GPS y generar continuamente retroalimentación a través del diálogo (e.g. entrenadores o atletas), haciendo que la información sea comprensible y el acuerdo sobre los puntos de acción (Gleason et al., 2023). Por último, tener un buen desarrollo de las habilidades sociales como ser afable y un comunicador eficaz, aspectos que son posiblemente más importantes que la generación de vastas bases de datos métricos.

Capítulo VIII. Referencias

- Adams, D., Morgans, R., Sacramento, J., Morgan, S., & Williams, M. D. (2013). Successful short passing frequency of defenders differentiates between top and bottom four English Premier League teams. *International Journal of Performance Analysis in Sport*, 13(3), 653-668. <https://doi.org/10.1080/24748668.2013.11868678>
- Ade, J., Fitzpatrick, J., & Bradley, P. S. (2016). High-intensity efforts in elite soccer matches and associated movement patterns, technical skills and tactical actions. Information for position-specific training drills. *J Sports Sci*, 34(24), 2205-2214. <https://doi.org/10.1080/02640414.2016.1217343>
- Akenhead, R., Harley, J. A., & Tweddle, S. P. (2016). Examining the External Training Load of an English Premier League Football Team With Special Reference to Acceleration. *J Strength Cond Res*, 30(9), 2424-2432. <https://doi.org/10.1519/jsc.0000000000001343>
- Almeida, C. H., Ferreira, A. P., & Volossovitch, A. (2014). Effects of Match Location, Match Status and Quality of Opposition on Regaining Possession in UEFA Champions League. *J Hum Kinet*, 41, 203-214. <https://doi.org/10.2478/hukin-2014-0048>
- Anaya Hernández, A. A. (2016). *Análisis y Lectura de Partidos*. México.
- Andrzejewski, M., Chmura, P., Konefał, M., Kowalczyk, E., & Chmura, J. (2018). Match outcome and sprinting activities in match play by elite German soccer players. *J Sports Med Phys Fitness*, 58(6), 785-792. <https://doi.org/10.23736/s0022-4707.17.07352-2>
- Aquino, R., Carling, C., Palucci Vieira, L. H., Martins, G., Jabor, G., Machado, J., Santiago, P., Garganta, J., & Puggina, E. (2020). Influence of Situational Variables, Team Formation, and Playing Position on Match Running Performance and Social Network Analysis in Brazilian Professional Soccer Players. *J Strength Cond Res*, 34(3), 808-817. <https://doi.org/10.1519/jsc.0000000000002725>
- Arjol-Serrano, J. L., Lampre, M., Díez, A., Castillo, D., Sanz-López, F., & Lozano, D. (2021). The Influence of Playing Formation on Physical Demands and Technical-Tactical Actions According to Playing Positions in an Elite Soccer Team. *Int J Environ Res Public Health*, 18(8). <https://doi.org/10.3390/ijerph18084148>
- Asian Clemente, J. A., Requena, B., Jukic, I., Nayler, J., Hernández, A. S., & Carling, C. (2019). Is Physical Performance a Differentiating Element between More or Less Successful Football Teams? *Sports (Basel)*, 7(10). <https://doi.org/10.3390/sports7100216>
- Barón, F., & Téllez, F. (2004). Apuntes de bioestadística. *Málaga: Universidad de Málaga*.
- Barron, D., Ball, G., Robins, M., & Sunderland, C. (2020). Identifying playing talent in professional football using artificial neural networks. *J Sports Sci*, 38(11-12), 1211-1220. <https://doi.org/10.1080/02640414.2019.1708036>
- Bastida Castillo, A., Gómez Carmona, C. D., De la Cruz Sánchez, E., & Pino Ortega, J. (2018). Accuracy, intra- and inter-unit reliability, and comparison between GPS and UWB-based position-tracking systems used for time-motion analyses in soccer. *Eur J Sport Sci*, 18(4), 450-457. <https://doi.org/10.1080/17461391.2018.1427796>
- Blei, D. M., & Smyth, P. (2017). Science and data science. *Proc Natl Acad Sci U S A*, 114(33), 8689-8692. <https://doi.org/10.1073/pnas.1702076114>

- Boada, A. J., & Mayorca, R. (2011). Planejamento de demanda, em empresas com estilo de venda por catálogo. *Revista lasallista de investigación*, 8(2), 124-135.
- Borghini, S., Colombo, D., La Torre, A., Banfi, G., Bonato, M., & Vitale, J. A. (2021). Differences in GPS variables according to playing formations and playing positions in U19 male soccer players. *Res Sports Med*, 29(3), 225-239. <https://doi.org/10.1080/15438627.2020.1815201>
- Bouza, C., & Santiago, A. (2012). La minería de datos: árboles de decisión y su aplicación en estudios médicos. *Modelación Matemática de Fenómenos del Medio Ambiente y la Salud*, 2, 64-78.
- Bradley, P. S., & Ade, J. D. (2018). Are Current Physical Match Performance Metrics in Elite Soccer Fit for Purpose or Is the Adoption of an Integrated Approach Needed? *Int J Sports Physiol Perform*, 13(5), 656-664. <https://doi.org/10.1123/ijsp.2017-0433>
- Bradley, P. S., Lago-Peñas, C., Rey, E., & Gomez Diaz, A. (2013). The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences*, 31(12), 1261-1270. <https://doi.org/10.1080/02640414.2013.786185>
- Bunker, R., & Thabtah, F. A. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15, 27-33. <https://doi.org/https://doi.org/10.1016/j.aci.2017.09.005>
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121-167.
- Bush, M., Barnes, C., Archer, D. T., Hogg, B., & Bradley, P. S. (2015). Evolution of match performance parameters for various playing positions in the English Premier League. *Hum Mov Sci*, 39, 1-11. <https://doi.org/10.1016/j.humov.2014.10.003>
- Carling, C. (2001). Getting the most out of football video and match analysis. *Insight FA Coaches Association Journal*, 5(3), 16-17.
- Carling, C., Williams, A. M., & Reilly, T. (2007). *Handbook of soccer match analysis: A systematic approach to improving performance*. Routledge.
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde, J., & Guerrero-Barrantes, M. (2016). Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Revista Tecnología en Marcha*, 29, 33-45.
- Casal, C. A., Andujar, M., Losada, J. L., Ardá, T., & Maneiro, R. (2016). Identification of Defensive Performance Factors in the 2010 FIFA World Cup South Africa. *Sports (Basel)*, 4(4). <https://doi.org/10.3390/sports4040054>
- Casal, C. A., Losada, J. L., Barreira, D., & Maneiro, R. (2021). Multivariate Exploratory Comparative Analysis of LaLiga Teams: Principal Component Analysis. *Int J Environ Res Public Health*, 18(6). <https://doi.org/10.3390/ijerph18063176>
- Castellano Paulis, J., & Casamichana Gómez, D. (2014a). Alternativas en la monitorización de las demandas físicas en fútbol: pasado, presente y futuro. *Revista Española de Educación Física y Deportes*(404), 41-58.
- Castellano Paulis, J., & Casamichana Gómez, D. (2014b). Deporte con dispositivos de posicionamiento global (GPS): Aplicaciones y limitaciones. *Revista de psicología del deporte*, 23(2), 0355-0364.

- Castellano Paulis, J., Casamichana Gómez, D., & Lago, C. (2012). The use of match statistics that discriminate between successful and unsuccessful soccer teams. *Journal of human kinetics*, *31*(2012), 137-147.
- Chmura, P., Andrzejewski, M., Konefał, M., Mroczek, D., Rokita, A., & Chmura, J. (2017). Analysis of Motor Activities of Professional Soccer Players during the 2014 World Cup in Brazil. *J Hum Kinet*, *56*, 187-195. <https://doi.org/10.1515/hukin-2017-0036>
- Chmura, P., Konefał, M., Chmura, J., Kowalczyk, E., Zając, T., Rokita, A., & Andrzejewski, M. (2018). Match outcome and running performance in different intensity ranges among elite soccer players. *Biology of sport*, *35*(2), 197-203.
- Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., & Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA),
- Claudino, J. G., Capanema, D. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Med Open*, *5*(1), 28. <https://doi.org/10.1186/s40798-019-0202-3>
- Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports Med*, *43*(10), 1025-1042. <https://doi.org/10.1007/s40279-013-0069-2>
- Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., Bisciotti, G. N., & Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European journal of sport science*, *11*(1), 51-59.
- Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., & Drust, B. (2009). Analysis of high intensity activity in Premier League soccer. *Int J Sports Med*, *30*(3), 205-212. <https://doi.org/10.1055/s-0028-1105950>
- Di Salvo, V., Pigozzi, F., González-Haro, C., Laughlin, M. S., & De Witt, J. K. (2013). Match performance comparison in top English soccer leagues. *Int J Sports Med*, *34*(6), 526-532. <https://doi.org/10.1055/s-0032-1327660>
- Dutky, S., & Maechler, M. (2022). *Package "bitops"*. <https://cran.r-project.org/web/packages/bitops/bitops.pdf>
- Eggels, H., Ruud, v. E., & Mycola, P. (2016). *Explaining Soccer Match Outcomes with Goal Scoring Opportunities Predictive Analytics* Proceedings of the 2016 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases,
- Eryarsoy, E., & Delen, D. (2019). *Predicting the Outcome of a Football Game: A Comparative Analysis of Single and Ensemble Analytics Methods* Proceedings of the 52nd Hawaii International Conference on System Sciences, <https://hdl.handle.net/10125/59550>
- Faude, O., Koch, T., & Meyer, T. (2012). Straight sprinting is the most frequent action in goal situations in professional football. *J Sports Sci*, *30*(7), 625-631. <https://doi.org/10.1080/02640414.2012.665940>
- Fernandez-Navarro, J., Fradua, L., Zubillaga, A., & McRobert, A. P. (2018). Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, *18*(3), 423-436.

- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., & Gorjanc, G. (2022). *Package "car"*. <https://cran.r-project.org/web/packages/car/car.pdf>
- Franks, I. M., & Miller, G. (1991). Training coaches to observe and remember. *J Sports Sci*, 9(3), 285-297. <https://doi.org/10.1080/02640419108729890>
- García-Aliaga, A., Marquina, M., Coterón, J., Rodríguez-González, A., & Luengo-Sánchez, S. (2021). In-game behaviour analysis of football players using machine learning techniques based on player statistics. *International Journal of Sports Science & Coaching*, 16(1), 148-157.
- García-Ramos, A., Haff, G. G., Feriche, B., & Jaric, S. (2018). Effects of different conditioning programmes on the performance of high-velocity soccer-related tasks: Systematic review and meta-analysis of controlled trials. *International Journal of Sports Science & Coaching*, 13(1), 129-151. <https://doi.org/10.1177/1747954117711096>
- Gervilla García, E., & Palmer Pol, A. L. (2009). Predicción del consumo de cocaína en adolescentes mediante árboles de decisión. *Revista de investigación en educación*, 6(1), 7-13.
- Geurkink, Y., Boone, J., Verstockt, S., & Bourgois, J. G. (2021). Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. *Applied Sciences*, 11(5), 2378.
- Gleason, B. H., Suchomel, T. J., Brewer, C., McMahon, E. L., Lis, R. P., & Stone, M. H. (2023). Defining the Sport Scientist. *Strength & Conditioning Journal*, 10.1519/SSC.0000000000000760. <https://doi.org/10.1519/ssc.0000000000000760>
- Groll, A., Ley, C., Schauburger, G., & Van Eetvelde, H. (2018). Prediction of the FIFA World Cup 2018-a random forest approach with an emphasis on estimated team ability parameters. *arXiv preprint arXiv:1806.03208*.
- Gu, W., Foster, K., Shang, J., & Wei, L. (2019). A game-predicting expert system using big data and machine learning. *Expert Systems with Applications*, 130, 293-305. <https://doi.org/https://doi.org/10.1016/j.eswa.2019.04.025>
- Hammer, B. (2003). *Mathematical Aspects of Neural Networks*. ESANN,
- Harper, D. J., Carling, C., & Kiely, J. (2019). High-Intensity Acceleration and Deceleration Demands in Elite Team Sports Competitive Match Play: A Systematic Review and Meta-Analysis of Observational Studies. *Sports Med*, 49(12), 1923-1947. <https://doi.org/10.1007/s40279-019-01170-1>
- Hassan, A., Akl, A. R., Hassan, I., & Sunderland, C. (2020). Predicting Wins, Losses and Attributes' Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis. *Sensors (Basel)*, 20(11). <https://doi.org/10.3390/s20113213>
- Hastie, T., Tibshirani, R., & Friedman, J. (2003). *The elements of statistical learning*. Springer.
- Hoppe, M., Slomka, M., Baumgart, C., Weber, H., & Freiwald, J. (2015). Match running performance and success across a season in German Bundesliga soccer teams. *International journal of sports medicine*, 36(07), 563-566.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., & Mitchell, D. (2022). *Package "lmtree"*. <https://cran.r-project.org/web/packages/lmtree/lmtree.pdf>
- Huang, K.-Y., & Chang, W.-L. (2010). *A neural network method for prediction of 2006 world cup football game* The 2010 international joint conference on neural networks (IJCNN), <https://doi.org/10.1109/IJCNN.2010.5596458>

- Hughes, M., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer-an analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sport and Exercise*, 7(2), 402-412. <https://doi.org/10.4100/jhse.2012.72.06>
- Hughes, M., & Franks, I. M. (2004). *Notational analysis—a review of the literature* (2nd ed.). Routledge. <https://doi.org/https://doi.org/10.4324/9780203641958>
- Ingebrigtsen, J., Dalen, T., Hjelde, G. H., Drust, B., & Wisløff, U. (2015). Acceleration and sprint profiles of a professional elite football team in match play. *Eur J Sport Sci*, 15(2), 101-110. <https://doi.org/10.1080/17461391.2014.933879>
- Isson, J. P. (2018). *Unstructured data analytics: How to improve customer acquisition, customer retention, and fraud detection and prevention*. John Wiley & Sons.
- Jamieson, J. P. (2010). The home field advantage in athletics: A meta-analysis. *Journal of Applied Social Psychology*, 40(7), 1819-1848.
- Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *Int J Sports Physiol Perform*, 13(5), 625-630. <https://doi.org/10.1123/ijsp.2017-0299>
- Kaplan, A. M., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business horizons*, 63(1), 37-50. <https://doi.org/10.1016/j.bushor.2019.09.003>
- Kempe, M., Vogelbein, M., & Nopp, S. (2016). The cream of the crop: Analysing FIFA world cup 2014 and Germany's title run. *Journal of Human Sport and Exercise*, 11(1), 42-52. <https://doi.org/https://doi.org/10.14198/jhse.2016.111.04>
- Konefał, M., Chmura, P., Kowalczyk, E., Figueiredo, A. J., Sarmiento, H., Rokita, A., Chmura, J., & Andrzejewski, M. (2019). Modeling of relationships between physical and technical activities and match outcome in elite German soccer players. *J Sports Med Phys Fitness*, 59(5), 752-759. <https://doi.org/https://doi.org/10.23736/S0022-4707.18.08506-7>
- Konefał, M., Chmura, P., Tessitore, A., Melcer, T., Kowalczyk, E., Chmura, J., & Andrzejewski, M. (2020). The impact of match location and players' physical and technical activities on winning in the German Bundesliga. *Frontiers in Psychology*, 11, 1748.
- Konefał, M., Chmura, P., Zajac, T., Chmura, J., Kowalczyk, E., & Andrzejewski, M. (2019a). Evolution of technical activity in various playing positions, in relation to match outcomes in professional soccer. *Biology of sport*, 36(2), 181-189.
- Konefał, M., Chmura, P., Zajac, T., Chmura, J., Kowalczyk, E., & Andrzejewski, M. (2019b). A New Approach to the Analysis of Pitch-Positions in Professional Soccer. *J Hum Kinet*, 66, 143-153. <https://doi.org/10.2478/hukin-2018-0067>
- Kubayi, A., & Toriola, A. (2020). Match Performance Indicators that Discriminated Between Winning, Drawing and Losing Teams in the 2017 AFCON Soccer Championship. *J Hum Kinet*, 72, 215-221. <https://doi.org/10.2478/hukin-2019-0108>
- Kuhn, M., & Wickham, H. (2022). *Package "tidymodels"*. <https://cran.r-project.org/web/packages/tidymodels/tidymodels.pdf>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2022). *Package "caret"*. <https://cran.r-project.org/web/packages/caret/caret.pdf>

- Lago-Peñas, C., Casais, L., Dominguez, E., & Sampaio, J. (2010). The effects of situational variables on distance covered at various speeds in elite soccer. *European journal of sport science*, *10*(2), 103-109.
- Lago-Peñas, C., & Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *Journal of human kinetics*, *25*(2010), 93-100. <https://doi.org/10.2478/v10078-010-0036-z>
- Lago-Peñas, C., Gómez-Ruano, M., & Yang, G. (2017). Styles of play in professional soccer: an approach of the Chinese Soccer Super League. *International Journal of Performance Analysis in Sport*, *17*(6), 1073-1084. <https://doi.org/10.1080/24748668.2018.1431857>
- Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of human kinetics*, *27*(1), 135-146.
- Link, D., & de Lorenzo, M. F. (2016). Seasonal Pacing - Match Importance Affects Activity in Professional Soccer. *PLoS One*, *11*(6), e0157127. <https://doi.org/10.1371/journal.pone.0157127>
- Liu, H., Hopkins, W. G., & Gómez, M. A. (2016). Modelling relationships between match events and match outcome in elite football. *Eur J Sport Sci*, *16*(5), 516-525. <https://doi.org/10.1080/17461391.2015.1042527>
- Longo, U. G., Sofi, F., Candela, V., Dinu, M., Cimmino, M., Massaroni, C., Schena, E., & Denaro, V. (2019). Performance activities and match outcomes of professional soccer teams during the 2016/2017 Serie A season. *Medicina*, *55*(8), 469. <https://doi.org/https://doi.org/10.3390/medicina55080469>
- López-Valenciano, A., Ruiz-Pérez, I., García-Gómez, A., Vera-García, F. J., De Ste Croix, M., Myer, G. D., & Ayala, F. (2020). Epidemiology of injuries in professional football: a systematic review and meta-analysis. *Br J Sports Med*, *54*(12), 711-718. <https://doi.org/10.1136/bjsports-2018-099577>
- Lord, F., Pyne, D. B., Welvaert, M., & Mara, J. K. (2020). Methods of performance analysis in team invasion sports: A systematic review. *J Sports Sci*, *38*(20), 2338-2349. <https://doi.org/10.1080/02640414.2020.1785185>
- Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: a critical review and implications for future research. *J Sports Sci*, *31*(6), 639-676. <https://doi.org/10.1080/02640414.2012.746720>
- Major League of Soccer. (2019). *2019 MLS Rosters*. <https://www.mlssoccer.com/rosters/2019>
- Malone, J. J., Barrett, S., Barnes, C., Twist, C., & Drust, B. (2020). To infinity and beyond: the use of GPS devices within the football codes. *Science and medicine in football*, *4*(1), 82-84.
- Malone, S., Owen, A., Mendes, B., Hughes, B., Collins, K., & Gabbett, T. J. (2018). High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? *J Sci Med Sport*, *21*(3), 257-262. <https://doi.org/10.1016/j.jsams.2017.05.016>
- Marcelino, R., Sampaio, J., & Mesquita, I. (2011). Investigação centrada na análise do jogo: da modelação estática à modelação dinâmica. *Revista Portuguesa de Ciências do Desporto*, *11*(1), 481-499.
- Mark, M. M., Mutrie, M., Brooks, D. R., & Harris, D. V. (1984). Causal Attributions of Winners and Losers in Individual Competitive Sports: Toward a Reformulation of

- the Self-Serving Bias. *Journal of Sport Psychology*, 6(2), 184-196. <https://doi.org/10.1123/jsp.6.2.184>
- Martinez, J. (2019). *Máquinas de Vectores de Soporte (SVM)*. <https://www.iartificial.net/maquinas-de-vectores-de-soporte-svm>
- Medina Merino, R. F., & Ñique Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*(10), 165-189. <https://doi.org/https://doi.org/10.26439/interfases2017.n10.1775>
- Memmert, D., Lemmink, K., & Sampaio, J. (2017). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Med*, 47(1), 1-10. <https://doi.org/10.1007/s40279-016-0562-5>
- Modric, T., Versic, S., & Jelcic, M. (2022). Monitoring technical performance in the UEFA Champions League: differences between successful and unsuccessful teams. *Montenegrin Journal of Sports Science and Medicine*, 11(2), 3-11.
- Modric, T., Versic, S., & Sekulic, D. (2020). Aerobic fitness and game performance indicators in professional football players; playing position specifics and associations. *Heliyon*, 6(11), e05427.
- Modric, T., Versic, S., Sekulic, D., & Liposek, S. (2019). Analysis of the Association between Running Performance and Game Performance Indicators in Professional Soccer Players. *International Journal of Environmental Research and Public Health*, 16(20), 4032. <https://www.mdpi.com/1660-4601/16/20/4032>
- Mohr, M., Krstrup, P., & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *J Sports Sci*, 21(7), 519-528. <https://doi.org/10.1080/0264041031000071182>
- Moncada-Jiménez, J., Salicetti-Fonseca, A., Carazo-Vargas, P., & Morera-Siércovich, P. L. (2021). La recolección, utilización y almacenamiento de datos biométricos sensibles en deportistas: insumos para la carrera de Educación Física. *Revista Educación*, 45(1), 640-652. <https://doi.org/10.15517/revedu.v45i1.41607>
- Moncada Jiménez, J., Carazo Vargas, P., Hernández Elizondo, J., Castillo Hernández, I., & Araya Vargas, G. (2023). *Metodología y estadística avanzada en las Ciencias del Movimiento Humano*. https://www.researchgate.net/publication/366920519_Metodologia_y_estadistica_avanzada_en_las_Ciencias_del_Movimiento_Humano
- Morera-Barrantes, R., Calderón-Chaverri, C., Gutiérrez-Vargas, R., Rojas-Valverde, D., Gutiérrez-Vargas, J. C., & Ugalde-Ramírez, J. A. (2021). Demandas físicas de jugadores profesionales costarricenses de fútbol: Influencia de la posición de juego y nivel competitivo. *MHSalud*, 18(2), 1-12.
- Muazu Musa, R., Razali Abdullah, M., Musawi Maliki, A. B. H., Azura Kosni, N., & Haque, M. (2016). The application of principal components analysis to recognize essential physical fitness components among youth development archers of Terengganu, Malaysia. *Indian Journal of Science and Technology*, 9(44), 1-6. <https://doi.org/10.17485/ijst/2016/v9i44/97045>
- Müller, K., Wickham, H., Francois, R., & Bryan, J. (2022). *Package "tibble"*. <https://cran.r-project.org/web/packages/tibble/tibble.pdf>
- Muñoz-López, A., Granero-Gil, P., Pino-Ortega, J., & De Hoyo, M. (2017). The validity and reliability of a 5-hz GPS device for quantifying athletes' sprints and movement demands specific to team sports. *Journal of Human Sport and Exercise*, 12(1), 156-166.

- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Muyor, J. M., Granero-Gil, P., & Pino-Ortega, J. (2018). Reliability and validity of a new accelerometer (Wimu®) system for measuring velocity during resistance exercises. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 232(3), 218-224.
- O'Donoghue, P. (2010). *Research methods for sports performance analysis*. Routledge.
- Oliva-Lozano, J. M., Fortes, V., Krstrup, P., & Muyor, J. M. (2020). Acceleration and sprint profiles of professional male football players in relation to playing position. *PLoS One*, 15(8), e0236959. <https://doi.org/10.1371/journal.pone.0236959>
- Oliva-Lozano, J. M., Fortes, V., & Muyor, J. M. (2023). When and how do elite soccer players sprint in match play? A longitudinal study in a professional soccer league. *Res Sports Med*, 31(1), 1-12. <https://doi.org/10.1080/15438627.2021.1929224>
- Oliva-Lozano, J. M., Rojas-Valverde, D., Gómez-Carmona, C. D., Fortes, V., & Pino-Ortega, J. (2021). Impact of contextual variables on the representative external load profile of Spanish professional soccer match-play: A full season study. *Eur J Sport Sci*, 21(4), 497-506. <https://doi.org/10.1080/17461391.2020.1751305>
- Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., & di Prampero, P. E. (2010). Energy cost and metabolic power in elite soccer: a new match analysis approach. *Medicine and science in sports and exercise*, 42(1), 170-178. <https://doi.org/10.1249/MSS.0b013e3181ae5cfd>
- Paredes-Inilupu, D. (2020). Data Science con R. Análisis de datos y algoritmos de predicción con R. In: Editorial Leanpub.
- Parmar, N., James, N., Hearne, G., & Jones, B. (2018). Using principal component analysis to develop performance indicators in professional rugby league. *International Journal of Performance Analysis in Sport*, 18(6), 938-949. <https://doi.org/10.1080/24748668.2018.1528525>
- Pedersen, T. L. (2022). *Package "patchwork"*. <https://cran.r-project.org/web/packages/patchwork/patchwork.pdf>
- Pino-Ortega, J., García-Rubio, J., & Ibáñez, S. J. (2018). Validity and reliability of the WIMU inertial device for the assessment of the vertical jump. *PeerJ*, 6, e4709.
- Pons, E., García-Calvo, T., Resta, R., Blanco, H., López Del Campo, R., Díaz García, J., & Pulido, J. J. (2019). A comparison of a GPS device and a multi-camera video technology during official soccer matches: Agreement between systems. *PLoS One*, 14(8), e0220729. <https://doi.org/10.1371/journal.pone.0220729>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rago, V., Rebelo, A., Krstrup, P., & Mohr, M. (2021). Contextual Variables and Training Load Throughout a Competitive Period in a Top-Level Male Soccer Team. *J Strength Cond Res*, 35(11), 3177-3183. <https://doi.org/10.1519/jsc.0000000000003258>
- Rampinini, E., Alberti, G., Fiorenza, M., Riggio, M., Sassi, R., Borges, T. O., & Coutts, A. J. (2015). Accuracy of GPS devices for measuring high-intensity running in field-based team sports. *Int J Sports Med*, 36(1), 49-53. <https://doi.org/10.1055/s-0034-1385866>
- Rampinini, E., Bishop, D., Marcora, S., Bravo, D. F., Sassi, R., & Impellizzeri, F. (2007). Validity of simple field tests as indicators of match-related physical performance in top-level professional soccer players. *International journal of sports medicine*, 28(03), 228-235.

- Rampinini, E., Coutts, A. J., Castagna, C., Sassi, R., & Impellizzeri, F. M. (2007). Variation in top level soccer match performance. *Int J Sports Med*, 28(12), 1018-1024. <https://doi.org/10.1055/s-2007-965158>
- Redwood-Brown, A., Cranton, W., & Sunderland, C. (2012). Validation of a real-time video analysis system for soccer. *Int J Sports Med*, 33(8), 635-640. <https://doi.org/10.1055/s-0032-1306326>
- Reep, C., & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4), 581-585.
- Reilly, T. P. (1975). *An ergonomic evaluation of occupational stress in professional football*. Liverpool John Moores University (United Kingdom).
- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *Springerplus*, 5(1), 1410. <https://doi.org/10.1186/s40064-016-3108-2>
- Rivas Borbón, O. M., Salas Cabrera, J., & Chávez Arce, T. S. (2017). Comparación del rendimiento físico de las selecciones nacionales de Alemania y Costa Rica, de acuerdo con los parámetros de metros recorridos en alta, mediana y baja intensidad y su relación con la posición alcanzada en la copa mundial de fútbol de Brasil 2014. *MHSalud*, 14(1), 1-13.
- Rizopoulos, D. (2015). *Package "bootStepAIC"*. <https://cran.r-project.org/web/packages/bootStepAIC/bootStepAIC.pdf>
- Rokach, L., & Maimon, O. (2007). *Introduction to decision trees* (2nd ed., Vol. 5). World Scientific Pub Co Inc. https://doi.org/https://doi.org/10.1142/9789812771728_0001
- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*, 13(7), e0201264. <https://doi.org/10.1371/journal.pone.0201264>
- Russell, M., Sparkes, W., Northeast, J., Cook, C. J., Love, T. D., Bracken, R. M., & Kilduff, L. P. (2016). Changes in Acceleration and Deceleration Capacity Throughout Professional Soccer Match-Play. *J Strength Cond Res*, 30(10), 2839-2844. <https://doi.org/10.1519/jsc.0000000000000805>
- Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). Match analysis in football: a systematic review. *J Sports Sci*, 32(20), 1831-1843. <https://doi.org/10.1080/02640414.2014.898852>
- Scott, M. T., Scott, T. J., & Kelly, V. G. (2016). The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *J Strength Cond Res*, 30(5), 1470-1490. <https://doi.org/10.1519/jsc.00000000000001221>
- Shi, L., Westerhuis, J. A., Rosén, J., Landberg, R., & Brunius, C. (2019). Variable selection and validation in multivariate modelling. *Bioinformatics*, 35(6), 972-980. <https://doi.org/10.1093/bioinformatics/bty710>
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., Despouy, P., Brüggemann, S., & Plotly Technologies Inc. (2022). *package "plotly"*. <https://cran.r-project.org/web/packages/plotly/plotly.pdf>
- Springham, M., Williams, S., Waldron, M., Strudwick, A. J., McLellan, C., & Newton, R. U. (2020). Prior workload has moderate effects on high-intensity match performance in elite-level professional football players when controlling for situational and contextual variables. *J Sports Sci*, 38(20), 2279-2290. <https://doi.org/10.1080/02640414.2020.1778355>

- Szwarc, A., Kromke, K., Radzimiński, Ł., & Jastrzębski, Z. (2017). Efficiency of 1-on-1 play situations for high-level soccer players during the World and European championships in relation to position on the pitch and match time. *International Journal of Sports Science & Coaching*, 12(4), 495-503. <https://doi.org/10.1177/1747954117717890>
- Taylor, J. B., Mellalieu, S. D., James, N., & Shearer, D. A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *J Sports Sci*, 26(9), 885-895. <https://doi.org/10.1080/02640410701836887>
- Therneau, T., Atkinson, B., & Ripley, B. (2022). *Package "rpart"*. <https://cran.r-project.org/web/packages/rpart/rpart.pdf>
- Tierney, P. J., Young, A., Clarke, N. D., & Duncan, M. J. (2016). Match play demands of 11 versus 11 professional football using Global Positioning System tracking: Variations across common playing formations. *Human movement science*, 49, 1-8. <https://doi.org/https://doi.org/10.1016/j.humov.2016.05.007>
- Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Med Open*, 8(1), 15. <https://doi.org/10.1186/s40798-022-00408-z>
- Ugalde Ramírez, J. A., Rivas Borbón, Ó. M., Gutiérrez Vargas, R., Rojas Valverde, D., & Sánchez Ureña, B. (2018). Parámetros Cinemáticos y Técnicos Realizados por los Futbolistas Según su Posición de Juego y su Relación con el Rendimiento Deportivo en el Mundial Brasil 2014. *Kronos: Rrevista universitaria de la actividad física y el deporte*, 17(2), 2.
- Van Gool, D., Van Gerven, D., & Boutmans, J. (1988). *The physiological load imposed on soccer players during real match-play*. E. & F.N. Spon.
- Varley, M. C., Jaspers, A., Helsen, W. F., & Malone, J. J. (2017). Methodological Considerations When Quantifying High-Intensity Efforts in Team Sport Using Global Positioning System Technology. *Int J Sports Physiol Perform*, 12(8), 1059-1068. <https://doi.org/10.1123/ijsp.2016-0534>
- Vigne, G., Gaudino, C., Rogowski, I., Alloatti, G., & Hautier, C. (2010). Activity profile in elite Italian soccer team. *Int J Sports Med*, 31(5), 304-310. <https://doi.org/10.1055/s-0030-1248320>
- Vogelbein, M., Nopp, S., & Hökelmann, A. (2014). Defensive transition in soccer - are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *J Sports Sci*, 32(11), 1076-1083. <https://doi.org/10.1080/02640414.2013.879671>
- Weaving, D., Jones, B., Ireton, M., Whitehead, S., Till, K., & Beggs, C. B. (2019). Overcoming the problem of multicollinearity in sports performance data: A novel application of partial least squares correlation analysis. *PLoS One*, 14(2), e0211776. <https://doi.org/10.1371/journal.pone.0211776>
- Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Package "dplyr"*. <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>
- Wickham, H., Vaughan, D., Girlich, M., Ushey, K., & Posit, P. (2021). *Package "Tidyr"*. <https://cran.r-project.org/web/packages/tidyr/tidyr.pdf>
- Williams, G., Vere Culp, M., Cox, E., Nolan, A., White, D., Medri, D., Waljee, A., Ripley, B., Magana, J., Tipparaju, S., Prasad Chappidi, D., Manyam Venkata, D.,

- Chakraborty, M., Zhou, F., & Chisholm, C. (2022). *Package "rattle"*. <https://cran.r-project.org/web/packages/rattle/rattle.pdf>
- Witte, T. H., & Wilson, A. M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *J Biomech*, 37(12), 1891-1898. <https://doi.org/10.1016/j.jbiomech.2004.02.031>
- Wright, M. N., Wager, S., & Probst, P. (2022). *Package "ranger"*. <https://cran.r-project.org/web/packages/ranger/ranger.pdf>
- Xie, Y., Cheng, J., Tan, X., Allaire, J. J., Girlich, M., Freedman Ellis, G., Rauh, J., SpryMedia Limited, Reavis, B., Gersen, L., Szopka, B., Pickering, A., Holmes, W., Marttila, M., Quintero, A., & Laurent, S. (2023). *Package "DT"*. <https://cran.r-project.org/web/packages/DT/DT.pdf>
- Zhou, C., Zhang, S., Lorenzo Calvo, A., & Cui, Y. (2018). Chinese soccer association super league, 2012–2017: key performance indicators in balance games. *International Journal of Performance Analysis in Sport*, 18(4), 645-656.

Anexo 1. Variables del WiMu Pro

Las variables que despliega el WIMU Pro se describen a continuación:

A. Variables relacionadas con distancia:

1. Distance(m): distancia recorrida en metros.
2. HSR Abs Dist (m): Abreviatura de “High Speed Running” absoluta, es la distancia recorrida a velocidades superiores a HIA (por defecto 21 km/h).
3. Vel Abs [12-19]: velocidad absoluta en zona.
4. Vel Abs [19-24]: velocidad absoluta en zona.
5. Vel Abs [24-50]: velocidad absoluta en zona.
6. Dist m/min: distancia relativa en metros por minuto

B. Variables relacionadas con la aceleración:

1. Acc Abs [3 10]: aceleración absoluta en la zona.
2. Acc Abs [-10 -3]: aceleración absoluta en la zona.

C. Variables relacionadas con los sprints o carreras cortas de alta intensidad:

1. Sprints ABS: número de sprints contados > 24 km/h.
2. MAX Speed(km/h): velocidad máxima alcanzada.

D. Variables relacionadas con la carga:

1. Player Load: Jugador/Carga de trabajo.
2. HMLD: Abreviatura de “High Metabolic Load Distance”, es la distancia en metros recorrida por un jugador cuando su potencia metabólica está por encima de 25.5 W/kg. Este valor de 25.5 corresponde a cuando el jugador está corriendo a una velocidad constante de 5.5 m/s² o cuando hay actividades de aceleración y desaceleración significativas.

Anexo 2. Variables del INSTAT Pro

Las variables que despliega el INSTAT Pro se describen a continuación:

- A. Estadísticas principales:**
 - 1. Índice InStat
 - 2. Partidos jugadores
 - 3. Minutos jugados
 - 4. Inicio equipo titular
 - 5. Jugador sustituido
 - 6. Posición
 - 7. Goles
 - 8. Asistencias
 - 9. Jugadas de gol
 - 10. Oportunidades de gol con éxito
 - 11. Ocasiones de gol, % conversión
 - 12. Ocasiones generadas
 - 13. Faltas
 - 14. Faltas recibidas
 - 15. Tarjetas amarillas
 - 16. Tarjetas rojas
 - 17. Fuera de juego
 - 18. Acciones totales
 - 19. Acciones efectivas
 - 20. % acciones efectivas
 - 21. Tiros
 - 22. Tiros a portería
 - 23. % de efectividad en lanzamientos
 - 24. Tiros por fuera
 - 25. Tiros interceptados
 - 26. Tiros al poste / palo

- B. Pases:**
 - 1. Pases
 - 2. Pases efectivos
 - 3. % de efectividad de pases
 - 4. Pases de finalización
 - 5. Pases de finalización efectivos
 - 6. Centros
 - 7. Centros efectivos
 - 8. % de efectividad de los centros

- C. Balones perdidos y recuperados:**
 - 1. Balones perdidos
 - 2. Pérdidas en campo propio
 - 3. Balones recuperados

4. Recuperaciones en campo rival

D. Disputas:

1. Disputas
2. Disputas ganadas
3. % disputas ganadas
4. Disputas defensivas
5. Disputas defensivas ganadas
6. % Disputas defensivas ganadas
7. Disputas en ataque
8. Disputas en ataque ganadas
9. % disputas en ataque ganadas
10. Disputas aéreas
11. Disputas por arriba ganadas
12. % disputas por arriba ganadas
13. Regates
14. Regates efectivos
15. Regates con éxito, %
16. Entradas
17. Entradas efectivas
18. Robos de balón con éxito, %
19. Interceptaciones
20. Rechaces

Definiciones:

Acc Abs (3 – 10 m/s): número total de aceleraciones entre 3 y 10 m/s

Acc Abs (-10 -03 m/s): número total de desaceleraciones entre -10 y -3 m/s

Acciones totales: es la cantidad total de todos los eventos, donde se suman, por ejemplo: todos los tipos de pases, de balones divididos, de interceptaciones, de recogidas, de circulación de pelota y mal control de pelota, los tiros a portería, los tiros rechazados y los goles encajados.

Acciones efectivas: número total de acciones realizadas con éxito. También disponible como porcentaje.

Acciones efectivas (%): porcentaje de acciones realizadas con éxito.

Asistencia: La última acción de un jugador del equipo autor del gol, antes de que éste sea marcado por un compañero, o un gol en propia meta.

Asistencias esperadas: Las asistencias esperadas miden la probabilidad de que un pase se convierta en una asistencia de gol. Se contabiliza el total de asistencias esperadas.

Balones perdidos: número total de balones perdidos.

Balones recuperados: número total de balones recuperados.

Centros: balón jugado desde las bandas ofensivas dirigido hacia un compañero en el área frente a la portería contraria.

Centros efectivos: número total de centros realizados con éxito.

Despeje: golpeo del balón, por parte de un jugador defensor orientado hacia zonas alejadas de la propia portería, sin un objetivo preciso o para un saque de banda/de esquina, jugando sobre seguro.

Disputas: todo tipo de disputas, que incluyen: luchas de balones divididos, disputas aéreas, regates, entradas, pérdidas tras una entrada del rival.

Disputas aéreas: un duelo aéreo se considera ganado a favor del jugador que toca primero el balón, pase lo que pase después. Un duelo aéreo que acabe en falta se considerará ganado a favor del jugador que haya sufrido la falta

Disputas ganadas (%): porcentaje de disputas ganadas

Disputas defensivas: cuando un jugador intenta desposeer a un jugador contrario para impedir el avance de un ataque.

Disputas defensivas ganadas: número total de disputas defensivas ganadas.

Disputas defensivas ganadas (%): porcentaje de disputas defensivas ganadas.

Disputas por arriba: disputa de dos rivales por el balón que se encuentra más arriba del nivel del hombro.

Disputas por arriba ganadas: número total de disputas por arriba ganadas.

Disputas por arriba ganadas (%): porcentaje de disputas por arriba ganadas.

Distancia (m): total de distancia recorrida en metros por un jugador durante el partido

Distancia por minuto: distancia recorrida en metros por minuto por un jugador durante el partido.

Efectividad de los centros (%): porcentaje de centros ejecutados con éxito.

Efectividad de pases (%): porcentaje de pases ejecutados con éxito

Efectividad en lanzamiento (%) porcentaje de lanzamiento acertado.

Entradas: acción del jugador que intenta quitarle el balón al jugador que lo posee. Del mismo modo, se llama entrada a la oposición al regate del rival.

Entradas efectivas: número total de entradas ejecutadas con éxito.

Faltas: infracción cometida por un jugador conforme a la regla 12 (1, 3) de las Reglas de Juego del IFAB.

Fueras de juego: un jugador está en posición de fuera de juego si interviene en una jugada y se encuentra más cerca de la línea de meta opuesta que la pelota y el penúltimo adversario. Se describe en la regla 11 de las Reglas de Juego del IFAB.

Gol: un gol marcado según lo especificado en la regla 10.1 de las Reglas de Juego del IFAB

Goles esperados: es un modelo predictivo ML utilizado para evaluar la probabilidad de anotar para cada tiro realizado en el juego

Carrera a alta intensidad o “*high-speed running distance*”: distancia recorrida a velocidad superior a 21km/h.

Intercepción: acción de recuperación del balón que interrumpe el ataque rival, generalmente después un tiro o un pase preciso efectuado por el equipo rival.

Lanzamiento: es una intención de pase que no parece tener un objetivo específico

Máxima velocidad (km/h): máxima velocidad alcanzada por un futbolista durante un partido

Oportunidad de gol: una oportunidad clara de marcar un gol. Predominantemente marcado en un tiro, lo que significa un tiro peligroso.

Oportunidades de gol con éxito: significa una ocasión de gol que se marcó.

Pase: acción de pasar el balón a un compañero de equipo. Es considerado largo cuando es mayor a 45 metros o un pase por alto de más de 25 metros.

Pase de finalización: pase al compañero en posición de gol.

Pases de finalización efectivos: pase al compañero en posición de gol que termina en anotación

Pases efectivos: cantidad de pases acertados

Pérdida del balón: acción del jugador atacante que, mediante un pase impreciso, regate fallido, etc, provoca el final de la posesión del balón del propio equipo. No se registrarán como pérdida si la posesión termina como producto de una falta del rival o de un tiro.

Pérdidas en campo propio: número total de pérdidas de balón en mitad de zona propia

Rechaces: recuperar la posesión del balón sin disputarlo, proveniente de un mal control o pase erróneo del rival.

Recuperación del balón: acción del jugador defensor que, mediante una interceptación, entrada, lucha, etc., pone fin a la posesión del equipo rival.

Recuperaciones en campo rival: número total de recuperaciones ejecutados en mitad de campo rival.

Regates: acción del jugador atacante con balón orientada a desbordar a su rival. Un regate efectivo supone simultáneamente una acción de entrada fallida para el jugador rival.

Regates efectivos: número total de regates ejecutados con éxito.

Regates con éxito (%): porcentaje de regates acertados.

Robos de balón con éxito (%): porcentaje de balones robados

Tarjetas amarillas: acción disciplinaria del árbitro que se indica mostrando una tarjeta amarilla conforme a la regla 12.3 de las Reglas de Juego del IFAB. Una segunda tarjeta amarilla se considera tarjeta roja.

Tarjeta roja: acción disciplinaria del árbitro que se indica mostrando una tarjeta roja conforme a la regla 12.3 de las Reglas de Juego del IFAB. Se muestra una única vez y excluye al infractor del partido. Las causas pueden ser faltas, mala conducta o el comportamiento antideportivo.

Tiros a portería: un intento hacia la portería contraria con la intención de marcar.

Tiros fuera: tiros realizados dirigidos por fuera de la portería

Tiros interceptados: remates que no lograron alcanzar la portería producto de un bloqueo

Velocidad Absoluta [12 - 19](m): distancia alcanzada durante carreras de velocidad entre 12 a 19 km/h.

Velocidad Absoluta [19 - 24](m): distancia alcanzada durante carreras de velocidad entre 19 a 24 km/h.


Velocidad o Sprint: distancia alcanzada durante carreras de velocidad mayores a 24 km/h

Anexo 3. Producción académica durante el programa de maestría

A. Publicaciones en revistas científicas

Morera-Siercovich, P. L., & Moncada-Jiménez, J. (2020). A Meta-Analysis of the Effect of Small-Sided Games on the Repeated Sprint Ability in Soccer. *International Journal of Physical Education, Fitness and Sports*, 9(4), 70-84. <https://doi.org/10.34256/ijpefs2048>


REVIEW ARTICLE



**International Journal of
PHYSICAL EDUCATION, FITNESS AND SPORTS**

DOI: 10.34256/ijpefs2048



REVIEW ARTICLE





**International Journal of
PHYSICAL EDUCATION, FITNESS AND SPORTS**

DOI: 10.34256/ijpefs2048

A Meta-Analysis of the Effect of Small-Sided Games on the Repeated Sprint Ability in Soccer

P. L. Morera-Siercovich ^a  and **J. Moncada-Jiménez** ^{a,b*} 

^a School of Physical Education and Sports, University of Costa Rica, Costa Rica
^b Human Movement Sciences Research Center (CIMOHU), University of Costa Rica, Costa Rica
 *Corresponding author Phone: +506 25112909. E-mail: jose.moncada@ucr.ac.cr
 DOI: <https://doi.org/10.34256/ijpefs2048>
 Received: 07-11-2020, Accepted: 15-12-2020; Published: 18-12-2020

Abstract: To determine the effect of small-sided games (SSG) on the repeated sprint ability (RSA) in male soccer players. A meta-analysis was conducted in which studies were selected from the electronic databases Medline, SPORTDiscus, Google Scholar, and PubMed. Results: Six studies were included and 44 effect sizes (ES) were computed from 116 men (age = 18.0 ± 2.8 yr., height = 175.8 ± 3.3 cm, weight = 68.6 ± 5.6 kg, body mass index [BMI] = 22.3 ± 1.1 kg·m⁻², VO₂max = 56.9 ± 2.1 ml·kg⁻¹·min⁻¹). The SSG training improved the overall RSA performance (ES = -.54, 95%CI = -.89, -.20, p < .05), reduced RSA total time (ES = -.41, 95%CI = -.81, -.01, p < .05), and fatigue index (ES = -.83, 95%CI = -1.65, -.02, p < 0.05). Moderator analysis showed that the fatigue index is impaired when several sessions/week are performed (β = .69, 95%CI = .29, 1.10, R² = 82.9, p = .001) and is enhanced when the duration of the intervention is extended over several weeks (β = -.25, 95%CI = -.47, -.03, R² = 56.6, p = .05). The methodological use of SSG enhances the capacity of repeated sprints, while technical and tactical elements are worked concomitantly.

Keywords: Soccer, Game, Training, Sprint, Performance.

Moncada Jiménez, J., Salicetti Fonseca, A., Carazo Vargas, P., & Morera Siercovich, P. L. (2021). La recolección, utilización y almacenamiento de datos biométricos sensibles en deportistas: insumos para la carrera de Educación Física. *Revista Educación*, 45(1), 1-11. <https://doi.org/https://doi.org/10.15517/revedu.v45i1.41607>

Ensayos


La recolección, utilización y almacenamiento de datos biométricos sensibles en deportistas: insumos para la carrera de Educación Física

Collection, Use and Storage of Sensitive Biometric Data for Athletes: Inputs for Physical Education Majors

José Moncada-Jiménez

Escuela de Educación Física y Deportes, Universidad de Costa Rica, Costa Rica

jose.moncada@ucr.ac.cr

 <https://orcid.org/0000-0001-9807-5163>


DOI: <https://doi.org/10.15517/revedu.v45i1.41607>

Redalyc: <http://www.redalyc.org/articulo.oa?id=44064134030>

Alejandro Salicetti-Fonseca

Costa Rica, Escuela de Educación Física y Deportes, Universidad de Costa Rica., Costa Rica

alejandro.salicetti@ucr.ac.cr

 <https://orcid.org/0000-0001-7605-1128>

Pedro Carazo-Vargas

Escuela de Educación Física y Deportes, Universidad de Costa Rica, Costa Rica

pedro.carazo@ucr.ac.cr

 <https://orcid.org/0000-0002-5374-7031>

Pier Luigi Morera-Siercovich

Escuela de Educación Física y Deportes, Universidad de Costa Rica, Costa Rica

psiercovich@yahoo.it

Recepción: 08 Mayo 2020
Aprobación: 02 Octubre 2020

RESUMEN:

Recientes avances tecnológicos, incluyendo espacios de almacenaje y algoritmos computacionales, permiten la recolección y almacenaje de una amplia cantidad de variables biométricas asociadas al rendimiento deportivo. El propósito de este ensayo es reconocer la importancia de la tecnología portátil en el ámbito del deporte, así como las leyes internacionales y nacionales que regulan la recolección, uso y almacenaje de datos biométricos de naturaleza sensible provenientes de deportistas, con un ejemplo específico para el caso de futbolistas profesionales costarricenses. La información presentada y discutida es un insumo relevante y actual para la formación de profesionales en Educación Física. A diferencia de la normativa internacional, se concluye que en Costa Rica no existe claridad acerca de la legislación relacionada con la biometría del deporte. También queda en evidencia que es necesaria la creación de programas educativos interdisciplinarios en analítica deportiva.

PALABRAS CLAVE: Regulaciones, Leyes, Dispositivos portátiles, Wearables, Deportes.