

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Assessment of the market value of Liga MX footballers based on their sporting performance and qualitative variables in the 2021-2022 season

Luis Fernando Cisneros Nuñez
Universidad de Monterrey (Mexico)
luis.cisneros@udem.edu

Miguel Angel García Cepeda
Universidad de Monterrey (Mexico)
miguela.garcia@udem.edu

Hernán Ayrton González Cruz
Universidad de Monterrey (Mexico)
hernan.gonzalezc@udem.edu

Álvaro Andrés Jasso Cantú
Universidad de Monterrey (Mexico)
alvaro.jasso@udem.edu

Jaime Lara Lara
Universidad de Monterrey (Mexico)
<https://orcid.org/0000-0001-8704-9824>
jaime.lara@udem.edu

David Alejandro Vélez González
Universidad de Monterrey (Mexico)
david.velez@udem.edu

RESUMEN

El presente estudio identifica las principales variables que influyen en el valor de mercado de los jugadores de fútbol de la Liga MX utilizando variables de desempeño en la temporada previa 2021-2022 y otras variables cualitativas. Utilizamos tres algoritmos para pronosticar el valor de mercado asignado por Transfermarkt: regresión lineal múltiple, árboles de decisión y bosques aleatorios. De acuerdo con los resultados de nuestro algoritmo con menor error cuadrático medio, los bosques aleatorios, las variables de desempeño que más influyen en el valor de mercado de un jugador son el total de acciones, los regates y los duelos en los que se participa. Además de las variables de desempeño, existen variables cualitativas muy relevantes, como el equipo, la vigencia de contrato y la participación en selección nacional. Se recomienda a los clubes de la Liga MX emplear métodos de análisis similares para utilizar su presupuesto de transferencias de jugadores de forma más eficiente.

PALABRAS CLAVE

Aprendizaje automático; valuación de futbolistas; Wyscout; Transfermarkt.

ABSTRACT

This work identifies the key variables influencing the market value of Liga MX soccer players, using performance information from the 2021-2022 season and other qualitative data. We employed three algorithms to predict the market value assigned by Transfermarkt: multiple linear regression, decision trees, and random forests. According to the model with the lower mean squared error (random forests), the performance variables with the greatest impact on a player's market value are total actions, dribbles, and duels in which the player participate. In addition to performance in the previous season, highly relevant qualitative variables include the player's team, contract year, and participation in the national team. We recommend that Liga MX clubs use similar analytical methods to allocate their financial resources more efficiently.

KEYWORDS

Machine learning; market value of football players; Wyscout; Transfermarkt.

Clasificación JEL: L8, M5, Z22, Z23.

MSC2010: 68T01, 62J05.

1. INTRODUCCIÓN

La asignación eficiente de recursos también ha sido estudiada por medio de la economía del deporte en el fútbol. Hogan y Massey (2021) analizaron los efectos de la compra realizada por Qatar Sports Investments del equipo francés de fútbol París Saint Germain (PSG). Dentro de sus hallazgos los autores determinaron que el PSG gastó 600 millones de euros en exceso para ganar la Liga Francesa entre las temporadas 2011/2012-2016/2017. De acuerdo con la teoría de los rendimientos marginales decrecientes cada punto adicional obtenido tras ganar la Liga es extremadamente costoso. Actualmente, autores como Al-Asadi y Tasmemir (2022), Behravan y Razavi (2021) Majewski (2016) y Metelski (2021) han estudiado la determinación del valor de mercado de los jugadores de fútbol; es decir, cuál es el monto al que los equipos de fútbol deberían comprar o vender a los jugadores.

El objetivo de este trabajo es identificar las principales variables que influyen en el valor de mercado de los jugadores de fútbol de la Liga MX y estimar los jugadores que podrían ser más

rentables de adquirir con base en sus habilidades futbolísticas. Esta liga es una de las más importantes fuera de Europa y Sudamérica, de acuerdo a la Federación Internacional de Historia y Estadística de Fútbol ocupaba el sitio 26 en 2022, por lo que su análisis aportará resultados que pueden compararse con literatura previa usualmente realizada en ligas mejor ubicadas en los rankings internacionales. Transfermarkt es una de las páginas más reconocidas en brindar estimaciones del valor de mercado de jugadores de fútbol profesional de casi todo el mundo. Para cumplir el objetivo de investigación haremos uso de modelos de regresión múltiple y dos algoritmos de aprendizaje automático empleados recientemente en la literatura: árboles de decisión y bosques aleatorios (Al-Asadi y Tasdemir, 2022). A diferencia de literatura previa se utilizarán variables explicativas de desempeño reales de los jugadores, a partir de la base de datos Wyscout.

Lo anterior, brindará una alternativa que permita optimizar recursos a un equipo de fútbol profesional y así replantear objetivos sin la necesidad de invertir capital excesivo en jugadores que estén valuados por encima de su valor real. Se recomienda que los equipos de fútbol implementen estas técnicas de valuación para hacer uso de su presupuesto de forma eficiente, fichando jugadores de calidad a un precio más accesible.

En la siguiente sección se hará una recopilación de investigaciones similares a la presente con el fin de identificar variables previamente encontradas como significativas. En la sección 3 se describirán las principales fuentes de datos de información y los algoritmos utilizados en esta investigación. Posteriormente, en la sección 4, se presentarán los resultados obtenidos y los supuestos detrás de ellos. Finalmente se procede con la discusión y conclusiones en la sección 5.

2. REVISIÓN DE LITERATURA

Los jugadores y sus contratos son los activos más relevantes para un equipo de fútbol (Franceschi et al., 2023). Por ello, según Metelski, (2021), un aspecto muy importante dentro del negocio del fútbol es la transferencia de jugadores. Asimismo, una de las decisiones más importantes de un directivo en torno a la transferencia de un jugador es determinar su precio de transferencia (Müller et al., 2017), para minimizar sus gastos y maximizar sus ingresos a través de la compra y venta de jugadores, respectivamente. Herm et al. (2014) definen el valor de mercado de un jugador como el monto estimado por el que su equipo lo podría vender a otro, mientras que el precio de transferencia es el monto verdaderamente pagado por un jugador en el mercado de transferencias. Lo anterior ha dado lugar a distintas investigaciones y propuestas sustentadas por modelos econométricos y análisis estadísticos con el fin de enriquecer la toma de decisiones respecto al mercado de transferencias.

Frick (2011) planteó diversos factores que influyen en el salario de un jugador como: los juegos en temporadas anteriores, apariciones internacionales y cantidad de goles. A su vez, Deutscher y Büschemann (2016) plantean otros aspectos que también influyen en el salario como la edad, experiencia, talento, posición de juego y país de origen. Lo anterior, aunado a otros factores no cuantitativos que influyen tanto en el salario como el valor de mercado de un jugador; inclusive, si se pudiera medir, el aspecto colectivo también sería un factor a considerar.

Uno de los problemas en torno a la información sobre el desempeño y salarios de los jugadores de fútbol es su disponibilidad. Por lo general esta información es privada. Por ejemplo, Simmons (2022), aseveró que en Norte América [EUA y Canadá] la disponibilidad de fuentes públicas de información sobre salarios de deportistas se ha mantenido, no obstante, en Europa estos recursos son severamente limitados. En este contexto, según Simmons, diversos investigadores se han visto obligados a utilizar medidas proxy sobre el salario de un jugador, por ejemplo, su valor estimado de mercado. Esto último con base en fuentes como Transfermarkt y Kicker.

En una revisión sistemática de la literatura acerca de los principales factores que explican el valor de mercado de un futbolista Franceschi et al. (2023) encuentran que las variables más utilizadas para explicar el valor de mercado son: edad, minutos jugados, goles, asistencias, tarjetas

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

amarillas y rojas, apariciones, pertenencia a la selección nacional y el pie con el que el jugador es diestro. De estas, las tarjetas y el pie con el que se es diestro tienden a no ser significativas en los estudios previos.

Existen otras variables cuantitativas y cualitativas que influyen en el valor de un jugador y la relevancia de estas variables depende en gran medida de su posición [defensa, mediocampista o delantero], inclusive la misma posición es un factor de importancia que puede perturbar su valor de mercado. Según Miao et al. (2015) los delanteros reciben muchos más premios y atención que los porteros, posiblemente debido a que el fútbol es un juego de ataque. De acuerdo con Sánchez et al., (2022) los delanteros y mediocampistas son más valorados en las 5 principales ligas europeas. Esto es congruente con el mercado de transferencias, en el que los jugadores más caros históricamente, son los delanteros (ver Tabla 1).

Tabla 1. Fichajes más caros de la historia del fútbol

Nombre del jugador	Posición	Fecha de Transferencia	Monto pagado por el jugador (millones de euros)
Neymar	Extremo Izquierdo	2017-2018	222.00
Kylian Mbappé	Centro Delantero	2018-2019	180.00
Phillipe Coutinho	Extremo Izquierdo	2017-2018	135.00
Joao Félix	Media Punta	2019-2020	127.20
Ousmane Dembélé	Extremo Derecho	2017-2018	125.00
Antoine Griezmann	Media Punta	2019-2020	120.00
Jack Grealish	Extremo Izquierdo	2021-2022	117.50
Cristiano Ronaldo	Centro Delantero	2018-2019	117.00
Eden Hazard	Extremo Izquierdo	2019-2020	115.00
Romelu Lukaku	Centro Delantero	2021-2022	113.00

Fuente: Elaboración propia con información de Transfermarkt.

Soto-Valero (2017) realizó una investigación en la que se identifican los atributos más importantes que debe tener un jugador (ver Tabla 2), mientras que Yaldo y Shamir (2017) concluyen que diferentes posiciones podrían requerir un conjunto de habilidades diferentes; por ende, deben ser separados según la posición dentro del terreno de juego para obtener resultados más consistentes.

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

Tabla 2. Atributos deseados para las diversas características de los jugadores según el videojuego EA Sports FIFA

Físicas	Edad	Defensivas	Marcaje
	Altura		Bloqueo de Pie
	Peso		Bloqueo con Barrida
	Masa Muscular		Agresividad
Ataque	Tiro Cruzado	Mentalidad	Intercepciones
	Precisión de Cabeceo		Posicionamiento
	Pases Cortos		Visión
	Voleas		Faltas
Movimiento	Aceleración	Portero	Lanzamientos
	Velocidad de Arranque		Manejo de manos
	Agilidad		Pateo de balón
	Reacción		Posicionamiento
Habilidades	Balance	Fuerza	Reflejos
	Drible		Potencia de Tiro
	Chanfle		Salto
	Ejecución de Tiro Libre		Energía
	Pases Largos		Fuerza
	Control de Balón		Pases Largos

Fuente: Adaptación de Soto-Valero 2017.

Por otro lado, hay factores no relacionados a las habilidades de un jugador que también influyen en su valor de mercado y/o salario, como patrocinadores, ventas de mercancías, derechos de imagen, entre otros (Yaldo & Shamir, 2017); inclusive es posible que la discriminación sea un factor que también influya en el salario de un deportista (Simmons, 2022) y, por lo tanto, en su valor de mercado.

A lo largo del tiempo las técnicas para determinar el valor de mercado teórico de un futbolista han aumentado, por lo que el valor monetario estimado de un jugador variará según el método que se emplee y las variables que se utilicen. En este contexto, Majewski, (2016) proporciona un análisis sobre los diferentes enfoques utilizados para cotizar jugadores (ver Tabla 3). Por su parte, Moya-Villanueva y Córdova-Ramos (2021) plantean que si todos los equipos de fútbol utilizaran la misma metodología para definir el precio de sus jugadores sus cotizaciones se alinearon de acuerdo a sus niveles de talento o destreza. Esto tal y como lo indica la teoría de los mercados eficientes.

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

Tabla 3. Tipología de evaluaciones del desempeño de los futbolistas.

Metodología	Método de expertos	Modelos econométricos	Métodos numéricos	Métodos con modelos de opciones	Enfoques mixtos
Datos	Desconocidos, basado en experiencias empíricas del experto	Datos deportivos, económicos, biológicos, valor hipotético de mercado	Valores hipotéticos de mercado (Transfermarkt)	Valores hipotéticos de mercado (Transfermarkt, índice de valores OPTA)	Toda la información disponible
Costos	Desconocidos (hipotéticamente altos)	Las principales fuentes de datos son gratuitas	Gratuitos	Los datos del índice de valores de valores OPTA no son gratuitos	Las principales fuentes de datos son gratuitas
Complejidad de Verificación	Difícil de verificar, desconocido.	Fácil de utilizar	Se comparan los valores de mercado Bajo nivel de complejidad		Fácil de utilizar.

Fuente: Adaptación de Metelski, 2021.

Los estudios previos han utilizado dos fuentes específicas para construir sus bases de datos, Transfermarkt y el videojuego FIFA. En la primera (Transfermarkt), diversos autores como Al-Asadi y Tasdemir (2022); Herm et al., (2014); Majewski (2016); Moya-Villanueva y Córdova-Ramos (2021), Sánchez et al. (2022) Stanojevic y Gyarmati (2016) y Yiğit et al., (2020) han recolectado el valor que el mercado proporciona a los jugadores de fútbol. Transfermarkt presenta sus estimaciones del valor teórico de mercado de jugadores de fútbol profesional de casi todo el mundo.

Por otra parte, autores como Awasthi et al. (2021); Behravan y Razavi, (2021), Shin y Gasparyan (2014), Singh y Lamba (2019), Soto-Valero (2017) y Yaldo y Shamir (2017) utilizaron las estadísticas proporcionadas por el videojuego FIFA para tener un parámetro en torno a las habilidades y características de los jugadores. Con el paso del tiempo los recursos disponibles para medir el desempeño futbolístico de un jugador se han ido acrecentando, por ejemplo: el análisis a través de video grabaciones (Shu-Ching Chen et al., 2005). El presente estudio utilizará información de los jugadores de la Liga MX recabada de la plataforma Wyscout y de Transfermarkt. Una de las ventajas de utilizar Wyscout es que proporciona estadísticas reales del desempeño obtenidas de las videograbaciones de los partidos de la temporada, información escasamente utilizada en la literatura (Hill et al., 2025; Pappalardo et al., 2019).

Asimismo, en los últimos años se ha empezado a utilizar modelos de aprendizaje automático para identificar las variables más influyentes en el valor de mercado de un futbolista y determinar ese mismo valor. De acuerdo con la investigación de Behravan y Razavi (2021) los métodos de regresión son los que destacan en las investigaciones en torno a los factores y/o variables que influyen en el valor de mercado de un futbolista, no obstante, las técnicas de aprendizaje automatizado han comenzado a tomar relevancia en este campo de estudio.

Al-Asadi y Tasdemir (2022) proponen un método cuantitativo para la determinación de mercado de los jugadores de futbol mediante el uso de la base de datos del FIFA 2020. En su investigación analizan los factores más importantes sobre la determinación de la valuación de mercado y se emplea el uso de regresiones lineales múltiples, árboles de decisión y bosques aleatorios. Encuentran que el algoritmo de bosques aleatorios tiene un mejor rendimiento en pronosticar el valor de mercado de los jugadores en comparación con la regresión múltiple, al reducir el error

de pronóstico en mayor medida que las regresiones múltiples habituales y otros algoritmos de aprendizaje automático. Ante esto, se optó por utilizar también los tres algoritmos en la presente investigación y observar cuál de ellos proporciona mejor ajuste.

3. METODOLOGÍA

3.1 Fuentes de Datos

Actualmente existen diversas fuentes de información como noticias, investigaciones y páginas web con valores estimados para la mayoría de los jugadores de fútbol profesional. Una de las plataformas más reconocidas a nivel mundial es Transfermarkt. La plataforma Transfermarkt ha sido estudiada por diversos autores como Metelski (2021), quien concluye que Transfermarkt es un buen proxy para definir el salario de un jugador, así como su valor de mercado. En adición, Prockl y Frick (2018) encontraron coeficientes de correlación de 0.70-0.75 entre el valor de mercado de un jugador [usando información de Transfermarkt] y su salario, esto para la Major Soccer League de EUA.

La plataforma de Transfermarkt provee un valor de mercado estimado a más de 800 mil jugadores. El valor de mercado estimado significa: el monto al que dos equipos deberían pactar una transacción de un jugador de fútbol. Transfermarkt realiza lo anterior mediante la metodología de Sabiduría de las Multitudes. De acuerdo al artículo de Peeters (2018), el valor de mercado viene determinado por los usuarios registrados de la plataforma que expresan sus opiniones sobre jugadores en específico, por lo que cada usuario expresa el valor estimado de cada jugador.

Christian Schwarz, jefe internacional de valores de mercado de Transfermarkt, comentó que la plataforma utiliza la 'Sabiduría de las Multitudes' y otros factores para realizar las valuaciones de los futbolistas. Primero, los usuarios de Transfermarkt (aproximadamente 1 millón) debaten y sugieren cuál debería de ser el valor de un jugador de fútbol. Posteriormente, cada liga de fútbol cuenta con un encargado voluntario que analiza los valores propuestos por los usuarios, además de ciertas variables para cada jugador como: los minutos jugados, goles, asistencias, partidos jugados con la selección, vigencia de contrato, salario, prestigio, y situación financiera del club (Keppel y Claessens, 2020). Realizado lo anterior, cada encargado de liga compara que los valores sean similares entre jugadores con el mismo nivel de habilidad. Finalmente, las valuaciones son enviadas a Christian Schwarz y él las aprueba o realiza ajustes en caso de ser necesario.

Muchos agentes de los jugadores actualmente utilizan las predicciones de Transfermarkt al momento de negociar futuros contratos para sus clientes (Stanojevic y Gyarmati, 2016). En esta investigación Transfermarkt es la plataforma en la que se obtuvo el valor estimado de los futbolistas de la Liga MX durante el mes de agosto de 2022.

Por su parte, las variables explicativas utilizadas en los modelos y algoritmos fueron obtenidas por medio de la compañía italiana Wyscout. Esta empresa italiana se encarga de realizar análisis de los partidos de fútbol y del rendimiento de los jugadores en dichos partidos. Su procedimiento de recolección de información es a través de un algoritmo entrenado para recolectar información sobre los jugadores a partir del video del juego. El algoritmo es supervisado por un experto; al terminar la grabación de cada partido, se efectúa un control de calidad reduciendo el margen de error, de tal manera que se evalúa con claridad y precisión la recopilación de los datos de los jugadores que el software registró. El operador brinda una doble verificación, por lo que al final, se obtiene un registro completo de las acciones que realizó el jugador en el partido. Se realiza a nivel de clubes y selección, recopilando en total más de cien variables por jugador (Pappalardo et al., 2019).

Las variables cualitativas obtenidas de la base de Wyscout fueron el nombre, equipo, nacionalidad, selección nacional, pie dominante y posición que desempeña dentro del campo de juego. Por el otro lado, las variables cuantitativas obtenidas de la misma manera de Wyscout fueron

edad, altura, peso, vigencia de contrato, minutos, acciones, acciones logradas, goles, asistencias, tiros, tiros exitosos, expectativa de gol, pases, pases exitosos, pases largos, pases largos exitosos, centros, centro acertados, regates, regates exitosos, duelos, duelos ganados, duelos aéreos, duelos aéreos ganados, intercepciones, balones perdidos, balones perdidos en propia mitad, balones recuperados, balones recuperados mitad avanzada, tarjetas amarillas y tarjetas rojas. Todas estas variables forman parte de las variables generales que se recopilan en general a todos los futbolistas en Wyscout. No se agregaron variables específicas por posición. Para el presente estudio, se realizó una recopilación de datos de los jugadores de los torneos de Apertura 2021 y Clausura 2022 de Liga MX en el mes de agosto del 2022.

Con el fin de tener una base de datos homogénea, solamente se utilizaron jugadores con más de 90 minutos de participación durante el año futbolístico ya mencionado. Esto debido a que la mayoría de los jugadores que cuentan con menos de 90 minutos de participación tienden a ser jugadores que acaban de debutar en primera división, por lo que no cuentan con una valuación de mercado en Transfermarkt.

La cantidad de jugadores recopilados para este estudio es variable dependiendo del equipo (algunos equipos pueden poseer 22-21 jugadores en su plantilla, cuando otros tienen 14-18). Esto ante los minutos que entraban al partido de cambio los jugadores suplentes y al mercado de fichajes en diciembre. Este último punto influye dado que ciertos jugadores cambiaron de equipo a mitad de temporada y en nuestra base de datos se les registró con el equipo con el que jugaron en la segunda mitad de la temporada (torneo de clausura). Adicionalmente, para la presente investigación se omitió la posición de los porteros debido a que es una posición en la que solo se registran como máximo a tres porteros por equipo y usualmente solo juega un portero por equipo en la temporada. Además, las variables cuantitativas que Wyscout registra para los porteros son distintas a las del resto de las posiciones del equipo.

Para explicar la influencia de la nacionalidad sobre el valor de mercado estimado de un jugador se creó una variable llamada confederación en la que se agrupó a los jugadores dependiendo de su país de nacimiento. En el caso de la Liga MX solo había jugadores de tres confederaciones, la Asociación del Norte, Centroamérica y el Caribe (CONCACAF), la Confederación Sudamericana de Fútbol (CONMEBOL) y la Unión de Federaciones Europeas de Fútbol (UEFA). Finalmente se generó una variable dicotómica con valor uno para los jugadores que fueron convocados por su respectiva selección durante la temporada 2021-2022 y vieron acción por lo menos en un partido.

3.2 Modelo de Regresión Lineal Múltiple

Un modelo de regresión lineal múltiple es una técnica estadística que usa un conjunto de variables independientes (x_1, x_2, \dots, x_p) con el objetivo de predecir el resultado de una variable dependiente (y). En otras palabras, usando este método se puede predecir una variable (valor de mercado), con base en otras variables que estén relacionadas, de acuerdo a la siguiente especificación

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + u$$

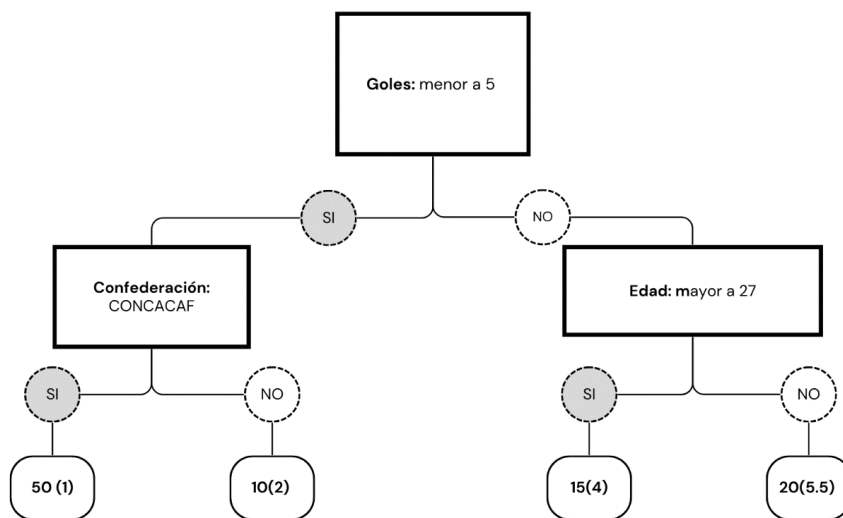
Aquí, u representa las variables no incluidas en el modelo que explican la variable dependiente y se conoce como el término de error. Los β_j se estiman utilizando los datos de una muestra por el método de mínimos cuadrados ordinarios. Para nuestro modelo de regresión lineal múltiple definiremos como nuestra variable dependiente (y) el precio de mercado según Transfermarkt, y como variables independientes (x_1, x_2, \dots, x_p) las listadas en la sección anterior. Para medir qué tanto los pronósticos del modelo se ajustan a los datos reales se puede utilizar la raíz del error cuadrático medio, también conocida como RMSE. Por cada observación se calcula la diferencia entre el valor real y el predicho, posterior a ello se eleva al cuadrado y se lleva a cabo una sumatoria de todos los valores. Esta suma de los errores al cuadrado se divide entre el número de observaciones ajustada por los grados de libertad y se obtiene la raíz. Mientras sea menor su valor en promedio el modelo se ajusta mejor al conjunto de datos.

3.3 Árboles de decisión

Un árbol de decisión es un método analítico que, mediante una representación esquemática de alternativas disponibles, permite la toma de decisiones de manera más sencilla y puede ser utilizado para pronosticar una variable continua como en la regresión múltiple. El árbol de decisión se representa a través de un mapa de los posibles resultados de las decisiones tomadas. En la Figura 1, a partir del nodo raíz, donde se encuentra toda la muestra, el algoritmo divide este nodo en dos nodos secundarios de tal manera que los valores de la variable de respuesta dentro de cada región secundaria sean lo más similares posible utilizando la suma de los cuadrados como criterio. Posteriormente, el proceso de división se repite para los nodos secundarios hasta que se satisface el criterio de detención, cuando el árbol ya acabó de ser construido.

En el ejemplo de la Figura 1 el nodo raíz se divide en dos nodos secundarios de acuerdo al número de goles entre aquellos que hayan anotado menos de 5 goles y los que hayan anotado 5 o más. El primer nodo secundario se divide de acuerdo a la pertenencia de la nacionalidad a la CONCACAF, y el segundo nodo secundario de acuerdo a la edad. Y así nos quedamos con cuatro regiones o nodos terminales. En esta Figura los jugadores que cuentan con una valuación más alta son aquellos que anotaron 5 goles o más y son menores de 27 años, en donde habría 20 jugadores. En cambio, los jugadores con una valuación más baja serían aquellos que anotaron menos de 5 goles y pertenecen a la CONCACAF, donde habría un total de 50 jugadores.

Figura 1. Ejemplo de Árbol de decisión



Fuente: Elaboración propia.

Los modelos estadísticos de aprendizaje automático pueden presentar casos de sobreajuste. Esto ocurre cuando el modelo obtenido en una muestra de entrenamiento se aplica a nueva información, dando como resultado que el modelo a la hora de realizar predicciones en el nuevo conjunto de información tiene un error alto. Por consiguiente, los modelos de aprendizaje automatizado, cuando sufren de sobreajuste, no son buenos para clasificar información nueva,

significando que el modelo pierde su capacidad generalizar y que, en vez de capturar patrones subyacentes, presta más atención a los detalles incluyendo los que no inciden de forma alguna sobre la predicción. Es por esta razón que un modelo predictivo de aprendizaje automático, se tiene que entrenar en un conjunto de datos y evaluar en otro llamado de prueba. En el primer conjunto de datos el algoritmo el modelo está aprendiendo, mientras que el conjunto de prueba sirve para evaluar el comportamiento del modelo ante nueva información (Ying, 2019). Para la presente investigación, se dividió a la muestra total en un 70 % como conjunto de entrenamiento y un 30 % como conjunto de prueba como habitualmente se hace en la literatura.

Una de las opciones para evitar problemas de sobreajuste es limitar el crecimiento del árbol. En la presente investigación, se utilizaron parámetros para obtener una cantidad mínima de observaciones por nodo, al igual que una cantidad mínima de observaciones finales por hoja en el árbol. Para que el algoritmo pueda generar una partición de los datos, necesita contar con un mínimo de 20 observaciones antes de crear un nuevo nodo. Al final de realizar las particiones, cada hoja final debía de contar con mínimo siete observaciones.

La importancia de cada variable en este algoritmo es estimada con el método de conteo, que cuenta cuantas veces una variable apareció como división principal en un nodo y como opción en las divisiones sustitutas, que sería el caso al no poder usarse la variable que mejor reduce el error de pronóstico, ponderando por la contribución a la reducción general en los errores y normalizando su importancia para que la suma de todas sea igual a 100. Una vez que se crea el árbol de decisión, el pronóstico se realiza clasificando a cada observación en una hoja final tras pasar por cada nodo del árbol y asignando como pronóstico el promedio de las observaciones en esa hoja en el conjunto de entrenamiento (Therneau y Atkinson, 2022). En nuestro ejemplo de la Figura 1, los 20 jugadores en la hoja final de valor más alto tienen en promedio 5.5 de valor.

3.4 Bosques Aleatorios

Los bosques aleatorios es otro algoritmo de aprendizaje automático, y puede ser utilizado para la regresión, siendo una alternativa para resolver los problemas de sobreajuste en los árboles de decisión. Surge como la agrupación de varios árboles de decisión. Para empezar, se requiere definir cuántos árboles hay en el bosque, en este caso para la presente investigación se utilizan 200 árboles. Posteriormente, se tiene que crear un subconjunto para cada uno de los árboles (con la data del conjunto de entrenamiento) introduciendo un componente aleatorio. Se tendrían que realizar 200 subconjuntos de entrenamiento para cada árbol. Para realizar cada subconjunto se emplea el muestreo al reemplazo o Bootstrapping tomando al azar observaciones (los jugadores) del conjunto de entrenamiento hasta obtener el número de observaciones necesarias para cada árbol. El término de reemplazo significa que una observación puede ser utilizada en múltiples ocasiones para completar un subconjunto de datos para cada árbol de decisión dentro del bosque. También selecciona de manera aleatoria una cantidad de variables con las cuales se hace cada partición de la muestra. Para efectos de pronóstico se realiza un promedio con el resultado de cada predicción de cada uno de los árboles dentro del bosque (LeDell, 2018). La importancia de las variables se calcula basado en la reducción de la suma de los errores cuadrados al momento de que esa variable es utilizada para dividir a un nodo (Kuhn, 2019).

4. RESULTADOS

4.1 Regresión múltiple

El modelo que pronostica el valor de mercado en millones de euros (MDE) utilizando todas las variables se encuentra en el Anexo A. De los resultados obtenidos, se infiere lo siguiente con las variables más significativas:

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

1. Respecto a la variable de equipo, jugar en un club distinto al América, Cruz Azul, León, Monterrey, Pachuca o Tigres afecta negativamente el valor pronosticado del jugador. En algunos casos esta diferencia es estadísticamente significativa con respecto al América.
2. Conforme la edad de los jugadores incrementa, el valor de mercado se ve afectado negativamente.
3. Ser convocado a su respectiva selección nacional genera un incremento en el valor de mercado. A su vez, pertenecer a otra confederación distinta a la CONCACAF (UEFA o CONMEBOL) eleva el valor de mercado.
4. Los jugadores con contratos más extensos tienen valuaciones más elevadas. En la presente investigación solamente es estadísticamente significativo tener contrato con vigencia hasta 2026. En este sentido, se infiere que los equipos buscan retener a los mejores jugadores el mayor tiempo posible. Aquellos jugadores con contrato a 2026 tienen un incremento en valor de 2.5 MDE con respecto a los jugadores con vigencia de contrato en 2020.
5. Contrario a lo esperado, los delanteros tienen una menor valuación con respecto a los medios.
6. De las variables de desempeño futbolístico, las más significativas son: goles, regates, y duelos aéreos exitosos. Las tres afectan de manera positiva al valor de mercado. También son relevantes los pases, los pases exitosos, los duelos exitosos (aunque con un signo diferente al esperado) y los duelos aéreos.

4.2 Árboles de Decisión

Al igual que con la regresión múltiple, al pronosticar mediante un árbol de decisión, se utilizó el conjunto de predictores que contiene todas las variables. El árbol obtenido se muestra en el Anexo B. Las variables más relevantes para el modelo, medidas por el método de conteo, son las que se encuentran en la Tabla 4.

Tabla 4. Variables más relevantes en el árbol de decisión medidas por el método de conteo

Equipo	Minutos	Acciones	Pases	Balones perdidos
11	10	10	9	8
Duelos	Intercepciones	Contrato	Edad	Expectativa de gol
7	6	5	5	3
Goles	Tiros	Balones Recuperados Mitad Avanzada	Selección Nacional	Acciones logradas
3	3	3	2	2

Fuente: Elaboración propia.

Nota: Se presentan las 15 variables más relevantes, entre mayor es el valor mayor es la importancia. El resto de variables tiene una importancia que sumaría 100.

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

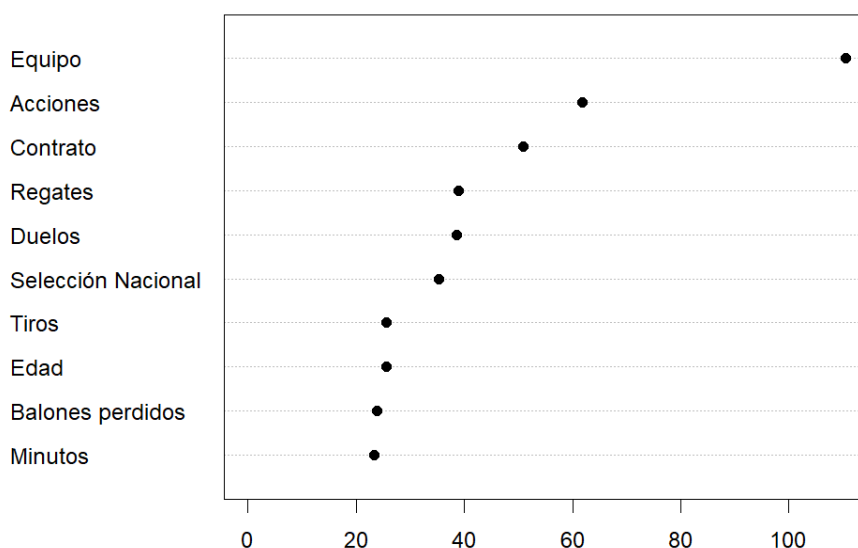
Las variables más relevantes son el equipo, los minutos, las acciones, los pases y los balones perdidos. Adicionalmente, los jugadores con los valores de mercado más elevados (7.3 MDE), son aquellos que:

1. Tuvieron más de 1,176 acciones en total durante la temporada.
2. No juegan en Atlético San Luis, Tijuana, Cruz Azul, Ciudad Juárez, Mazatlán, Necaxa, Puebla, UNAM, Querétaro y Toluca. La mayoría de estos equipos coinciden con los que suelen estar en la parte baja de la obtención de puntos en la temporada.
3. Son menores de 30 años.
4. Tienen un contrato vigente hasta 2021 o 2026.

4.3 Bosques aleatorios

Al realizar los pronósticos de los bosques aleatorios se utilizaron las mismas variables que en los algoritmos anteriores. Las variables más relevantes son medidas por la reducción media en la precisión se pueden observar en la Figura 2.

Figura 2. Importancia de Variables en Bosques Aleatorios.



Fuente: Elaboración propia.

Nota: En el eje horizontal se presenta la reducción media en la precisión (Khun, 2019)

Para el caso de la presente investigación, el impacto que tiene el equipo sobre la certeza de la predicción del valor de mercado es el más elevado. En adición, otras variables cualitativas como la selección nacional y el contrato forman parte de las variables más determinantes. Dentro de las variables de desempeño, el número total de acciones, los regates, los duelos, los tiros, los balones perdidos y los minutos son los que más explican el valor de mercado de los futbolistas. La edad también tiene un valor relevante.

4.4 Análisis de Algoritmos

Todos los algoritmos fueron entrenados con un conjunto de entrenamiento y posteriormente fueron evaluados en el conjunto de prueba. Se calcularon los errores cuadráticos medios (RMSE) de cada modelo con el conjunto de prueba (ver Tabla 5). Se concluyó que bosques aleatorios es el algoritmo que mejor pronostica el valor de mercado de un jugador de la Liga MX al tener un menor error cuadrático medio, de forma similar a los resultados de Al-Asadi y Tasdemir (2022).

Tabla 5. Raíz de los errores cuadráticos medios de cada algoritmo.

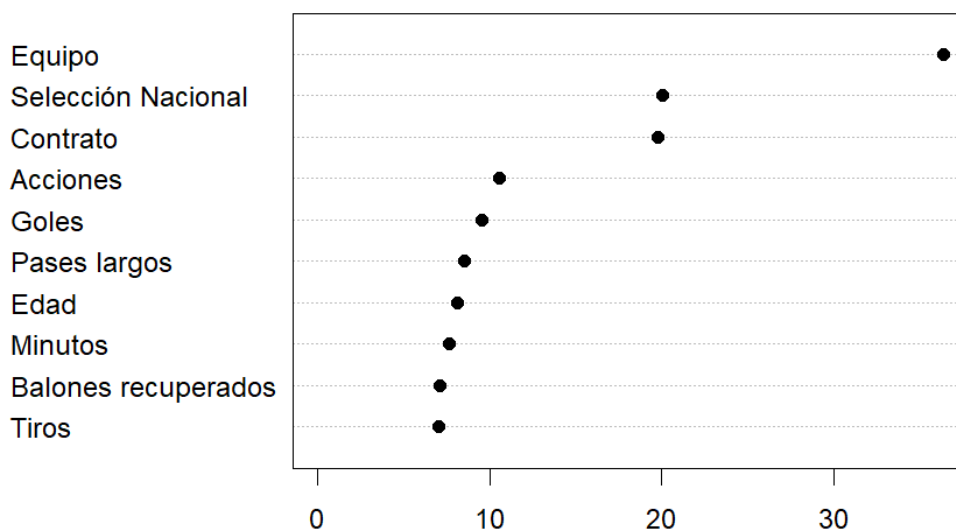
Modelo	RMSE
Regresión Lineal Múltiple	1.66
Árbol de Decisión	1.80
Bosques Aleatorios	1.43

Fuente: Elaboración propia.

4.5 Análisis por Posición

Tras concluir que el algoritmo de bosques aleatorios es el más efectivo para pronosticar en la presente investigación, se volvió a estimar el mismo algoritmo, pero en tres muestras diferentes por posición de juego. Esto con el fin de determinar si existen variaciones en la importancia de las variables a diferencia del modelo para todas las posiciones. Los resultados se muestran en las Figuras 3-5.

Figura 3. Importancia de Variables en Bosques Aleatorios por Defensas.



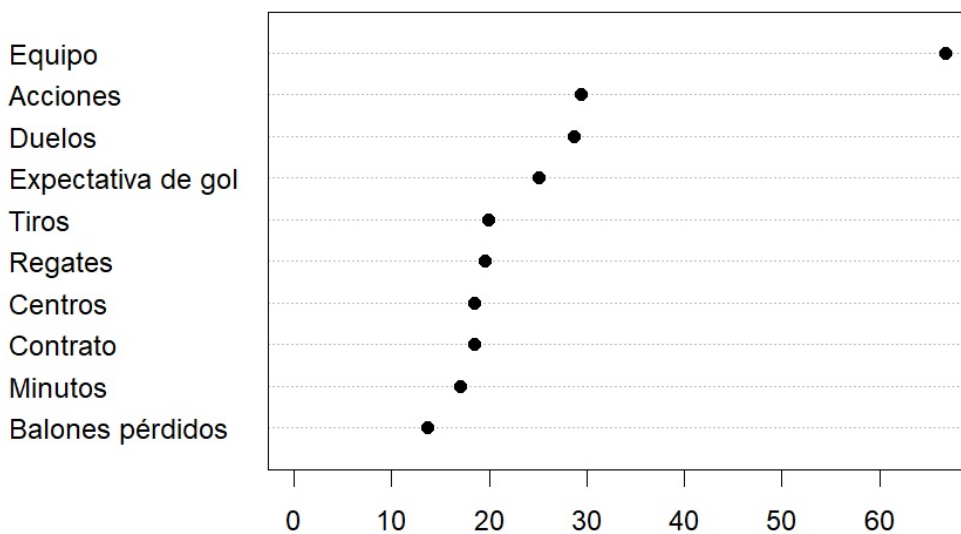
Fuente: Elaboración propia.

Nota: En el eje horizontal se presenta la reducción media en la precisión (Khun, 2019)

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

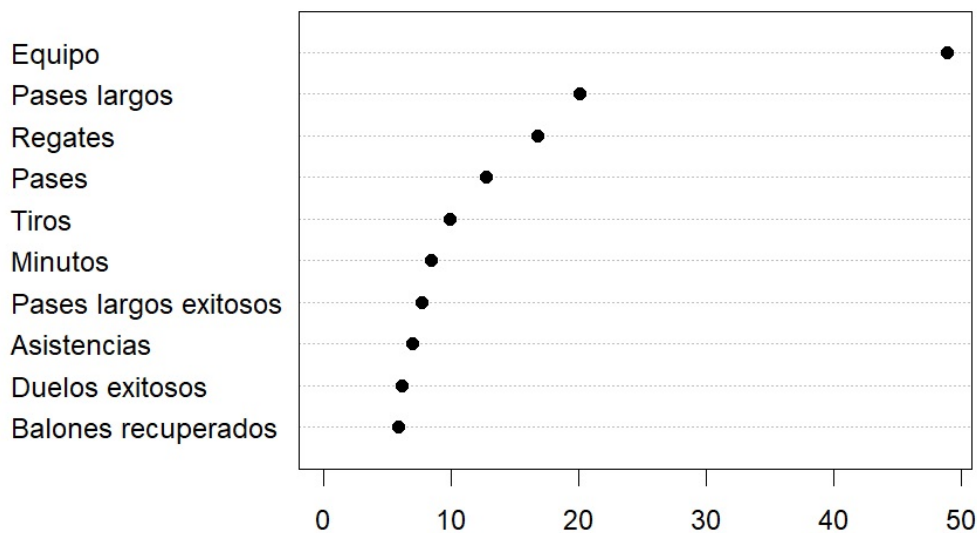
Figura 4. Importancia de Variables en Bosques Aleatorios por Medios



Fuente: Elaboración propia.

Nota: En el eje horizontal se presenta la reducción media en la precisión (Khun, 2019)

Figura 5. Importancia de Variables en Bosques Aleatorios por Delanteros.



Fuente: Elaboración propia.

Nota: En el eje horizontal se presenta la reducción media en la precisión (Khun, 2019)

Con los resultados obtenidos, las variables relevantes para la posición de cada jugador son distintas. Sin embargo, el equipo sigue apareciendo como la variable más relevante en todas las posiciones. En el caso de los defensas otras dos variables cualitativas le siguen: selección nacional y la vigencia del contrato. Aparecen algunas variables relacionadas con su incorporación en tareas ofensivas como los goles y los pases largos. Mientras que la variable más relevante sobre sus funciones que aparece es la de balones recuperados. En el caso de los centrocampistas, después del equipo, las acciones totales y los duelos en que participan destacan algunas relacionadas con tareas ofensivas como expectativa de gol, tiros, regates y centros. Por el otro lado, los balones perdidos parecen afectar de manera negativa al valor de mercado. Por último, en el modelo para los delanteros, sobresalen también algunas variables de desempeño ofensivo como los pases largos, los regates, los pases, los tiros y las asistencias. Asimismo, los duelos exitosos y pases largos exitosos sobresalen en el desempeño de esta función; es posible que, a diferencia de otras posiciones, la precisión de estas acciones es más relevante en los delanteros ya que ocurre en un espacio de la cancha donde es más difícil acertar. De forma un tanto extraña, no aparece la variable de goles. También, para los delanteros destaca una variable de habilidades defensivas como el recuperar balones. En general, la relevancia de las variables de los jugadores por posición de juego, es congruente con los roles generales que cada posición de juego debe cumplir dentro del campo, aunque parece dársele un poco de mayor relevancia a las variables ofensivas en todas las posiciones. El algoritmo tuvo un menor error cuadrático en el caso de los defensas.

5. DISCUSIÓN Y CONCLUSIONES

Se logró identificar las principales variables que influyen en el valor de mercado de los jugadores de fútbol de la Liga MX. Se puede concluir que existen variables distintas al rendimiento deportivo dentro de la cancha que integran el valor de mercado, tal como se mide en la plataforma Transfermarkt. Se realizaron análisis de datos utilizando tres algoritmos distintos, la regresión lineal múltiple, árboles de decisión y bosques aleatorios, para determinar las variables principales que afectan el valor de mercado de un futbolista. Con los tres algoritmos se obtuvieron pocas semejanzas respecto a las variables más significativas e importantes; el equipo al que cada jugador pertenece, los llamados a la selección nacional, el contrato, la edad y los regates aparecen en los tres algoritmos.

Nuestros resultados coinciden en algunos aspectos con la literatura previa. Con respecto a las variables más utilizadas en la literatura revisada por Franceschi et al. (2023) encontramos que la edad, la pertenencia a la selección nacional y variables de desempeño deportivo suelen ser significativas. Al mismo tiempo variables como las tarjetas o el pie con el que un futbolista es diestro no resultan significativas. De forma relevante y novedosa encontramos que el equipo al que se pertenece en la Liga MX tiene una correlación significativa con los valores en Transfermarkt. Nuestros resultados también coinciden en que la utilización de algoritmos de aprendizaje automatizado puede reducir el error de pronóstico del valor de mercado que se obtiene con regresiones múltiples como en Al-Asadi y Tasdemir (2022), particularmente se coincide en que bosques aleatorios produce el menor error cuadrático medio; además, las variables relevantes pueden diferir en distintos algoritmos.

Actualmente, con los avances tecnológicos que miden el desempeño de todos los jugadores en cada partido, se abren nuevas puertas para la investigación de la economía del deporte en el fútbol. Estos avances permiten el uso de bases de datos reales del rendimiento de un jugador, como Wyscout, que había sido empleada escasamente en la literatura previa (Hill et al., 2025).

Al comparar los resultados de los modelos de aprendizaje automatizado se encuentran varias semejanzas. En ambos modelos (árboles de decisión y bosques aleatorios) las acciones y el equipo son las dos principales variables explicativas. Se infiere que aquellos jugadores que participan más en los partidos, cuentan con valores más elevados. Adicionalmente, la vigencia de contrato es una variable relevante en ambos modelos.

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

De las estimaciones realizadas utilizando el modelo de bosques aleatorios, el cual tiene un menor error cuadrático medio, se obtienen los siguientes resultados puntuales especiales:

1. El jugador más infravalorado es Avilés Hurtado de Pachuca que está valuado en 1.2 MDE en Transfermarkt y de acuerdo a nuestro modelo su valor de mercado debería de ser de 4.21 MDE. Las principales razones por las cuales nuestro modelo pronostica un valor más elevado es dado que juega en el tercer equipo con más puntos en el torneo y se encuentra dentro del 25% de los jugadores con más minutos jugados y más acciones en el torneo.
2. El jugador más sobrevalorado es Carlos Alberto Rodríguez de Cruz Azul, cuyo valor en Transfermarkt (10 MDE) se encuentra 6.17 MDE por encima de la estimación obtenida del modelo utilizado.
3. El jugador más valioso de acuerdo a nuestro modelo es el francés Florian Thauvin de Tigres, cuyo valor pronosticado es de 8.11 MDE. El valor de Transfermarkt de Thauvin es el más caro de la Liga MX y es de 12 MDE.

Las principales limitaciones del estudio giran en torno a que únicamente se consideran las variables futbolísticas de la temporada 2021-2022, por lo que, si un jugador tuvo una lesión o una mala temporada, la estimación de su valor de mercado probablemente sea menor a lo que debería ante la trayectoria previa al análisis; ya que no se toman en cuenta sus estadísticas de temporadas anteriores. De la misma manera, la experiencia de los jugadores únicamente es medida por su convocatoria a la selección, por el club al que pertenecen y por el país donde nacieron, sin tomar en cuenta en qué países y clubes ha jugado ni qué premios o títulos ha conseguido en su carrera. Al no medir la experiencia previa, es posible que el equipo en el que el jugador este contratado funcione como una variable proxy del desempeño en temporadas previas, aunque quizá también la relevancia del equipo pueda deberse a que los participantes en Transfermarkt sesgan las valuaciones hacia equipos más conocidos o de mejor desempeño reciente. Lo mismo podría ocurrir con los llamados a la selección nacional o la vigencia del contrato.

Para futuras investigaciones se recomienda añadir variables para medir la experiencia de cada jugador de maneras adicionales. Algunas de estas podrían la inclusión de más temporadas previas, el club en donde se formó el jugador, de que equipo anterior proviene y si ha jugado en Europa. De la misma manera para contar con una base de datos más amplia, se recomienda ir agregando a otras Ligas de fútbol para ver qué variables son más significativas en otros países, de modo que podamos establecer si las variables de desempeño en la cancha son más relevantes en otras ligas. Nuestro trabajo solamente exploró la utilización de dos de los algoritmos más conocidos de aprendizaje automatizado. Dada la posibilidad de mejorar los pronósticos del valor de mercado con este tipo de algoritmos se sugiere analizar el desempeño de otros algoritmos dentro de esta literatura.

REFERENCIAS

- Al-Asadi, M. A., & Tasdemir, S. (2022). Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. *IEEE Access*, 10, 22631–22645. <https://doi.org/10.1109/ACCESS.2022.3154767>
- Awasthi, P., Beutel, A., Kleindessner, M., Morgenstern, J., & Wang, X. (2021). *Evaluating Fairness of Machine Learning Models Under Uncertain and Incomplete Information*. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 206–214. <https://doi.org/10.1145/3442188.3445884>
- Behravan, I., & Razavi, S. M. (2021). A Novel Machine Learning Method for Estimating Football Players' Value in the Transfer Market. *Soft Computing*, 25(3), 2499–2511. <https://doi.org/10.1007/s00500-020-05319-3>
- Deutscher, C., & Büschemann, A. (2016). Does Performance Consistency Pay Off Financially for Players? Evidence From the Bundesliga. *Journal of Sports Economics*, 17(1), 27–43. <https://doi.org/10.1177/1527002514521428>
- Franceschi, M., Brocard, J., Follert, F., & Gouguet, J. (2023). Determinants of Football Players' Valuation: A Systematic Review. *Journal of Economic Surveys*, 38(3), 577–600. <https://doi.org/10.1111/joes.12552>
- Frick, B. (2011). Performance, Salaries and Contract Length: Empirical Evidence from German Soccer. *International Journal of Sport Finance*, 6(2), 87–118.
- Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the Crowd Evaluates Soccer players' Market Values: Accuracy and Evaluation Attributes of an Online Community. *Sport Management Review*, 17(4), 484–492. <https://doi.org/10.1016/j.smr.2013.12.006>
- Hill, D. F., Skinner, J., & Grosman, A. (2025). A Review of Football Player Metrics and Valuation Methods: A Typological Framework of Football Player Valuations. *Managing Sport and Leisure*, 1–24. <https://doi.org/10.1080/23750472.2025.2459727>
- Hogan, V., & Massey, P. (2021). Soccer Clubs and Diminishing Returns: The Case of Paris Saint-Germain. *International Journal of Sport Finance*, 16(3), 127–138. <https://doi.org/10.32731/IJSF/163.082021.02>
- Keppel, P., & Claessens, T. (2020). How the Volunteers of Data Website Transfermarkt Became Influential Players at European top Football Clubs. *Follow the Money*. <https://www.ftm.eu/articles/transfermarkt-volunteers-european-football>
- Kuhn, M. (2019). *The Caret Package*. <https://topepo.github.io/caret/index.html>
- LeDell, E. (2018). *user! Machine Learning Tutorial*. <https://koalaverse.github.io/machine-learning-in-R/>
- Majewski, S. (2016). Identification of Factors Determining Market Value of the Most Valuable Football Players. *Journal of Management and Business Administration. Central Europe*, 24(3), 91–104. <https://doi.org/10.7206/jmba.ce.2450-7814.177>
- Metelski, A. (2021). Factors Affecting the Value of Football Players in the Transfer Market. *Journal of Physical Education and Sport*, 21(2), 1150 – 1155. <https://doi.org/10.7752/jpes.2021.s2145>
- Miao, H., Cachucho, R., & Knobbe, A. (2015). *Football Player's Performance and Market Value*. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD).
- Moya-Villanueva, G. R., & Córdova-Ramos, L. A. (2021). Análisis del Precio de los Futbolistas de la Liga Premier Inglesa con Respecto a sus Estadísticas en FIFA 20. *Pensamiento Crítico*, 26(2), 165–200. <https://doi.org/10.15381/pc.v26i2.18785>
- Müller, O., Simons, A., & Weinmann, M. (2017). Beyond Crowd Judgments: Data-Driven Estimation of Market Value in Association Football. *European Journal of Operational Research*, 263(2), 611–624. <https://doi.org/10.1016/j.ejor.2017.05.005>
- Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A Public Data set of Spatio-Temporal Match Events in Soccer Competitions. *Scientific Data*, 6(1), 236. <https://doi.org/10.1038/s41597-019-0247-7>
- Peeters, T. (2018). Testing the Wisdom of Crowds in the Field: Transfermarkt Valuations and International Soccer Results. *International Journal of Forecasting*, 34(1), 17–29. <https://doi.org/10.1016/j.ijforecast.2017.08.002>
- Prockl, F., & Frick, B. (2018). *Wage Determination in a Regulated Labor Market: Empirical Evidence from*

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

Major League Soccer (No. 39). Paderborn University, Faculty of Business Administration and Economics.

Sánchez, M., Orgaz, B., Ramirez-Campillo, R., Nakamura, F. Y., Luis-Pereira, J. M., Carretero, G., & Sánchez-Sánchez, J. (2022). Factores Asociados al Valor de Mercado de los Futbolistas Profesionales. *Cuadernos de Psicología Del Deporte*, 22(3), 140–150.

Shin, J., & Gasparyan, R. (2014). *A Novel Way to Soccer Match Prediction*. Stanford University: Department of Computer Science.

Shu-Ching Chen, Mei-Ling Shyu, & Na Zhao. (2005). *An Enhanced Query Model for Soccer Video Retrieval Using Temporal Relationships*. 21st International Conference on Data Engineering (ICDE'05), 1133–1134. <https://doi.org/10.1109/ICDE.2005.20>

Simmons, R. (2022). Professional Labor Markets in the Journal of Sports Economics. *Journal of Sports Economics*, 23(6), 728–748. <https://doi.org/10.1177/15270025211051062>

Singh, P., & Lamba, P. S. (2019). Influence of Crowdsourcing, Popularity and Previous Year Statistics in Market Value Estimation of Football Players. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(2), 113–126. <https://doi.org/10.1080/09720529.2019.1576333>

Soto-Valero, C. (2017). A Gaussian Mixture Clustering Model for Characterizing Football Players Using the EA Sports' FIFA Video Games System. [Modelo Basado en Agrupamiento de Mixturas Gaussianas para Caracterizar Futbolistas Utilizando el Sistema de Videojuegos FIFA de EA Sports. RICYDE. *Revista Internacional de Ciencias Del Deporte*, 13(49), 244–259. <https://doi.org/10.5232/ricyde2017.04904>

Stanojevic, R., & Gyarmati, L. (2016). *Towards Data-Driven Football Player Assessment*. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 167–172. <https://doi.org/10.1109/ICDMW.2016.0031>

Therneau, T. M., & Atkinson, E. J. (2022). *An Introduction to Recursive Partitioning Using the RPART Routines*. <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>

Yaldo, L., & Shamir, L. (2017). Computational Estimation of Football Player Wages. *International Journal of Computer Science in Sport*, 16(1), 18–38. <https://doi.org/10.1515/ijcss-2017-0002>

Yiğit, A. T., Samak, B., & Kaya, T. (2020). *Football Player Value Assessment Using Machine Learning Techniques*. International Conference on Intelligent and Fuzzy Systems, 289–297. https://doi.org/10.1007/978-3-030-23756-1_36

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168, 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

ANEXOS

Anexo A. Estimadores del Modelo de Regresión Lineal

	Estimador	Error Estándar	Nivel de Significancia
Constante	7.809	3.369	*
Atlas	-1.147	0.531	*
Atlético de San Luis	-0.827	0.494	.
Club Tijuana	-1.000	0.566	.
Cruz Azul	0.098	0.523	
Guadalajara	-0.113	0.516	
Juárez	-0.507	0.463	
León	0.030	0.640	
Mazatlán	-0.570	0.585	
Monterrey	1.123	0.548	*
Necaxa	-0.839	0.512	
Pachuca	0.485	0.585	
Puebla	-0.839	0.469	.
Pumas	-0.652	0.545	
Querétaro	-1.166	0.513	*
Santos	-0.248	0.506	
Tigres	0.355	0.482	
Toluca	-0.568	0.476	
Edad	-0.121	0.026	***
Selección nacional	0.796	0.240	**
CONMEBOL	0.941	0.218	***
UEFA	1.020	0.365	**
Altura	-0.031	0.021	
Peso	0.009	0.020	
Contrato 2021	0.701	0.411	.
Contrato 2022	0.368	0.387	
Contrato 2023	0.233	0.397	

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

	Estimador	Error Estándar	Nivel de Significancia
Contrato 2024	0.530	0.417	
Contrato 2025	0.172	0.463	
Contrato 2026	2.525	0.641	***
Perfil Izquierdo	-0.020	0.201	
Defensa	0.087	0.276	
Delantero	-0.627	0.304	*
Minutos	-0.001	0.000	
Acciones	0.000	0.001	
Acciones logradas	-0.009	0.011	
Goles	0.194	0.072	**
Asistencias	0.086	0.068	
Tiros	0.004	0.012	
Tiros exitosos	0.000	0.005	
Expectativa De Gol	-0.088	0.065	
Pases	0.002	0.001	.
Pases exitosos	0.014	0.008	.
Pases Largos	0.005	0.004	
Pases Largos exitosos	0.002	0.006	
Centros	0.003	0.005	
Centros exitosos	-0.003	0.004	
Regates	0.014	0.005	**
Regates exitosos	-0.001	0.005	
Duelos	-0.002	0.002	
Duelos ganados	-0.023	0.012	.
Duelos Aéreos	0.007	0.004	.
Duelos Aéreos ganados	0.018	0.007	**
Intercepciones	-0.010	0.007	
Balones perdidos	0.001	0.005	
Balones perdidos en propia mitad	0.008	0.007	

Evaluación del valor de mercado de los futbolistas de la Liga MX con base en su desempeño deportivo y variables cualitativas en la temporada 2021-2022

Luis Fernando Cisneros Nuñez; Miguel Angel García Cepeda; Hernán Ayrton González Cruz; Álvaro Andrés Jasso Cantú; Jaime Lara Lara; David Alejandro Vélez González

	Estimador	Error Estándar	Nivel de Significancia
Balones recuperados	0.006	0.004	
Balones recuperados mitad avanzada	0.008	0.008	
Tarjetas amarillas	-0.024	0.039	
Tarjetas rojas	0.119	0.159	

Nivel de Significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '' 0.1 ' ' 1. Para los efectos por equipo el club base es el América. En las confederaciones la CONCACAF. En las posiciones los medios. Para el contrato el año base es 2020.

ANEXO B: ÁRBOL DE DECISIÓN

