

Modelado Estadístico del Rendimiento Deportivo: Factores que Explican la Probabilidad de Victoria de Tigres 2024-2025

Statistical Modeling of Sports Performance: Factors that Explain the Probability of Victory of Tigres in 2024-2025

José Carlos Espinoza Briones¹

Autor de correspondencia: José Carlos Espinoza Briones, jose.espinozabr@uanl.edu.mx

¹Universidad Autónoma de Nuevo León.

Cómo citar:

Cómo citar: Espinoza, J. C. (2025). Modelado Estadístico del Rendimiento Deportivo: Factores que Explican la Probabilidad de Victoria de Tigres 2024-2025. *Revista De Ciencias Del Ejercicio FOD*, 20(2), 71-84.

DOI: 10.29105/rce-fod.v20i2.150

Julio-Diciembre-115

Link para acceder al artículo:

<https://doi.org/10.29105/rcefod.v19i1.115>



Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de Creative Commons Licencia de atribución (CC BY-NC) (Creative Commons Atribución-No-Comercial 4.0)

Resumen

El análisis estadístico es una herramienta clave para la toma de decisiones en el fútbol profesional. Este estudio utiliza un modelo logit para determinar la probabilidad de que el equipo de fútbol Tigres obtenga una victoria durante la temporada 2024-2025, a partir de variables asociadas al rendimiento deportivo del equipo. Se encontró que el factor más determinante para lograr el triunfo es la localía, mientras que, obtener mayor número de toques de balón, entradas al rival y tarjetas amarillas, en comparación con el rival, se asocia positivamente con la probabilidad de victoria del equipo. Por el contrario, alcanzar un mayor número de centros al área que el rival, se relaciona negativamente con dicha probabilidad. Por lo tanto, la presente investigación funge como elemento de análisis empírico para concretar una mejor toma de decisiones estratégicas, tanto para una directiva como para un cuerpo técnico, además, representa un caso de estadística aplicada en el deporte, lo cual puede ser útil para la docencia, la investigación y la formación de profesionales en ciencias del ejercicio, de la actividad física y del deporte.

Palabras Clave: fútbol; Liga MX; regresión logística; análisis predictivo; rendimiento competitivo.

Abstract

Statistical analysis is a key tool for decision-making in professional soccer. This study uses a logit model to estimate the probability of the Tigres soccer team winning during the 2024-2025 season, based on variables associated with the team's performance. It was found that the most decisive factor for achieving victory is home advantage, while obtaining a greater number of touches, tackles, and yellow cards compared to the opponent is positively associated with the team's probability of victory. Conversely, achieving a greater number of crosses into the box than the opponent is negatively associated with this probability. Therefore, this research provides an empirical basis to facilitate better strategic decision-making for both management and coaching staff. It also represents a case of statistics applied to sports, which can be useful for teaching, research, and the training of professionals in exercise science, physical activity, and sports.

Key words: soccer; Liga MX; logistic regression; predictive analytics; competitive performance.

Introducción

Si bien, la estadística descriptiva permite conocer el comportamiento de una variable respecto a una muestra o incluso a una población, representa solo el primer paso dentro del análisis estadístico. La estadística, en su sentido más amplio, tiene como objetivo realizar inferencias sobre una población a partir de los datos obtenidos de una muestra, incorporando medidas de certeza asociadas al proceso inferencial (Wackerly et al., 2008). Así, la estadística inferencial hace posible identificar relaciones entre variables y generalizar los resultados, siempre con un grado de certeza expresado en términos probabilísticos. En el ámbito deportivo, resulta especialmente relevante fortalecer el uso de la estadística inferencial, no solo aumentando su presencia, sino también mejorando la calidad de su aplicación.

Este enfoque científico en el deporte no es reciente. Desde 1950, el entrenamiento deportivo ha incorporado el método científico, particularmente a través del desarrollo de la periodización tradicional propuesta por el ruso Lev Pavlovich Matveyev, considerado el padre de la teoría del entrenamiento deportivo (Marques Junior, 2024). En su modelo, Matveyev (1977) identificó cinco componentes fundamentales de la preparación del deportista: física, técnica, táctica, psicológica y teórica. Todos ellos orientados hacia un objetivo central: ganar la competencia. En consecuencia, los resultados obtenidos en las competencias se convierten en el principal indicador del éxito de dicha preparación.

En éste sentido, el análisis del rendimiento en el fútbol ha evolucionado hacia una comprensión más integral del juego, en donde las acciones observables durante un partido reflejan la aplicación de principios tácticos, físicos, técnicos y psicológicos.

Según Tamarit (2007), en su obra titulada "¿Qué es la periodización táctica?", cada acción del juego debe analizarse en función de su vinculación con el modelo de juego del equipo, ya que el rendimiento no puede entenderse aislado de la estructura colectiva que lo sostiene. Es así que, variables como la posesión de balón, las intercepciones, las entradas o los duelos ganados no son simplemente estadísticas, sino manifestaciones cuantificables del dominio de ciertos principios de juego, como la organización defensiva, la transición o la eficacia ofensiva. Estas métricas, por tanto, pueden ser interpretadas como indicadores indirectos del cumplimiento del modelo táctico propuesto por el cuerpo técnico, y en consecuencia, del éxito competitivo. Esta visión teórica fundamenta el uso de dichas variables para modelar probabilísticamente el resultado de un partido. Es por ello que en éste estudio, la temática principal será el modelado matemático de la victoria de un club de fútbol a partir de los indicadores de rendimiento deportivo durante la competencia. Si bien, el tema de la presente investigación no es un tópico recurrente en los estudios científicos del deporte ni en los estudios de estadística aplicada, existen diversas investigaciones a nivel internacional que han abordado dicha temática.

Tal es el caso de Martínez Martínez y González García (2019), quienes analizaron las cinco grandes ligas de fútbol (España, Inglaterra, Italia, Alemania y Francia) en la temporada 2016-2017, y concluyeron que marcar gol primero que el rival, incrementa las probabilidades de obtener un triunfo al final del partido, y dichas probabilidades son mayores en el caso de hacerlo jugando de local.

Por su parte, Lepschy et al. (2020) condujeron un estudio que abordó la identificación de factores de triunfo en la liga de Alemania en el periodo 2014-2017, donde se encontró que los errores defensivos, el valor de mercado, la eficacia goleadora, los tiros a contraataque, los tiros a

puerta y el total de tiros tienen el mayor impacto positivo sobre un triunfo, mientras que los centros mostraron una relación negativa.

Desde su perspectiva, Elyakim et al. (2020) examinaron más de 455 parámetros de juego en 1,240 partidos de las ligas de fútbol de Italia e Israel, y determinaron que los pases clave, tiros a puerta y precisión en el pase, fueron los elementos de mayor importancia para ganar un partido en ambas ligas. Similarmente, Chocano Feito y Castilla González (2021) estudiaron los principales determinantes para predecir un resultado de fútbol en las temporadas 2016-2017 y 2017-2018 en Alemania, Italia, Inglaterra y España, a través de estadística multivariante, en particular, análisis de componentes principales y regresión logística, donde los disparos a favor y en contra son los dos factores más importantes de obtener una victoria, en contraparte, los tiros de esquina a favor de un equipo se asocian con menores probabilidades de triunfo.

De igual manera, Arrunategui Zeta (2021) analizó la liga peruana de fútbol en la temporada 2015-2019, donde ha encontrado que las tarjetas rojas y el sobrepeso de los jugadores disminuyen las probabilidades de ganar o empatar un partido, por el contrario, la localía, mayor posesión de balón y contar con jugadores más altos y de mayor edad que el promedio, incrementan dichas probabilidades. En este sentido, Fernández-Cortés Tolosa et al. (2022) examinaron las temporadas 2014-2015, 2015-2016 y 2016-2017 de la liga española, y encontraron que, los equipos que anotan el primer gol, independientemente de la condición de localía, consiguen un mayor número de victorias.

Por su parte, Botana García (2022) determinó mediante el uso regresiones logísticas y modelos de árboles de decisión, que la posesión de balón no es significativa, sin interacciones con otras variables, en las probabilidades de que un equipo de fútbol de las cinco grandes ligas obtenga una victoria. Mientras que, Ovando Fuentealba (2025) empleó modelos estadísticos de aprendizaje automático (machine learning) como random forest, XGBoost, CatBoost y regresión logística para predecir una victoria local, una victoria o un empate, en el fútbol chileno y sostiene que el empate es el resultado más complejo de predecir.

Sin embargo, en la actualidad no existen estudios estadísticos aplicados para el caso de México. Es por ello, que la presente investigación puede fungir como elemento exploratorio y como un punto de partida importante en la materia. De hecho, a pesar de los avances tecnológicos y la creciente disponibilidad de datos sobre rendimiento deportivo, persiste una falta de estudios que analicen el impacto de variables tácticas y físicas sobre la probabilidad de victoria en el fútbol profesional mexicano.

La realización de este estudio aporta un análisis estadístico inferencial sobre el rendimiento deportivo de un equipo del fútbol mexicano, generando información valiosa para los distintos agentes. Para los directivos, ofrece una base objetiva para tomar decisiones estratégicas relacionadas con la gestión deportiva, fichajes o alineaciones. Para el cuerpo técnico, proporciona una herramienta de retroalimentación sobre los indicadores que se asocian a la probabilidad de victoria, lo que puede orientar entrenamientos y tácticas. Para los aficionados, permite una comprensión más profunda del juego y de los factores que contribuyen al éxito del equipo.

Finalmente, para el ámbito académico, el estudio representa un caso aplicado de análisis de datos en el deporte, útil para la docencia, la investigación y la formación de profesionales en ciencias del ejercicio, ciencias de la actividad física, ciencias del deporte, estadística aplicada y gestión deportiva. En conjunto, estos aportes fortalecen el puente entre el análisis cuantitativo y la práctica futbolística en el contexto nacional.

Por todo lo anterior, el presente estudio tiene como objetivo determinar la probabilidad de que el equipo de Tigres obtenga una victoria durante la temporada 2024-2025, empleando un modelo logit con variables asociadas al rendimiento deportivo del equipo. Por lo que se plantea la siguiente pregunta de investigación: ¿cómo influyen las variables de rendimiento deportivo en la probabilidad de triunfo en el equipo de Tigres durante la temporada 2024-2025? Se plantea la hipótesis de que existe una relación estadísticamente significativa entre las variables de rendimiento deportivo y la probabilidad estimada de victoria, modelada mediante regresión logística.

Metodología

Diseño

El presente es un estudio cuantitativo, de tipo predictivo y no experimental, ya que se basa en la observación de variables sin manipulación de las mismas. Se utiliza un modelo logit (regresión logística binaria) para estimar la probabilidad de que el equipo Tigres gane un partido durante la temporada 2024-2025, en función de variables relacionadas con el rendimiento deportivo.

Muestra

La muestra está compuesta por los partidos disputados por el equipo de Tigres durante la temporada regular 2024-2025, que incluye los torneos Apertura 2024 y Clausura 2025. En total, se consideran 34 observaciones, correspondientes a los 34 partidos que el equipo juega (17 por torneo). Cada observación representa un partido individual. Si bien, no se tiene un gran número de observaciones, se cuenta con al menos 10 eventos por predictor como se realiza en la práctica con fines de estabilidad y precisión, ya que el equipo resultó ganador en 20 encuentros, de los restantes 14, empató y perdió el mismo número de encuentros, siete. Por otro lado, el modelo es exploratorio, no para predicción robusta, sino para generar hipótesis o identificar patrones preliminares, por lo que, contar con 34 observaciones limita el poder predictivo pero no lo invalida, lo cual se trata de una limitación del estudio.

Dado que el objetivo del modelo es generar predicciones relevantes y aplicables en el contexto competitivo actual, se considera adecuado utilizar información exclusivamente reciente. Incluir temporadas anteriores podría inducir sesgos derivados de condiciones pasadas que ya no se mantienen como cambios de reglas o normativas, distintos jugadores, entrenadores y directiva. Por lo que, el estudio se enfoca exclusivamente en la temporada 2024-2025 debido a que esta ofrece una unidad temporal homogénea en cuanto a contexto competitivo. Esta delimitación permite realizar un análisis específico, pertinente y aplicable al entorno actual del equipo, además de cumplir con los requisitos mínimos metodológicos para la estimación del modelo logit.

Instrumentos

Los datos utilizados provienen del sitio especializado en estadística deportiva Sports-Reference (2025), que recopila información detallada sobre los partidos de fútbol de la liga mexicana. Las variables extraídas incluyen estadísticas como posesión de balón, tarjetas, faltas, tiros de esquina, centros, entradas, intercepciones, entre otros. No se utilizó ningún instrumento físico de recolección, ya que los datos fueron tomados de una base digital confiable, datos que se descargaron directamente del sitio una vez concluido el año futbolístico, sitio que es de acceso público.

Las variables que serán consideradas inicialmente en el modelo son todas aquéllas contenidas en el sitio especializado en estadística deportiva Sports-Reference (2025), relacionadas al rendimiento deportivo del equipo.

Figura 1

Variables de rendimiento deportivo que serán consideradas en el modelo

Local	Posesión de balón	Tarjetas amarillas	Tarjetas rojas	Faltas	Tiros de esquina	Centros	Toques de balón
Entradas	Intercepciones	Duelos aéreos ganados	Despejes	Fueras de lugar	Saques de meta	Saques de banda	Balones largos

La Figura 1 contiene las posibles variables a considerar en el modelo, donde local es una variable binaria, toma el valor de 1 cuando Tigres es local y 0 cuando es visitante. Posesión de balón se refiere al porcentaje de tiempo que un equipo controla el balón durante el partido. Tarjetas amarillas representa el número de sanciones disciplinarias por alguna infracción moderada. Tarjetas rojas es el número de sanciones debidas a una falta grave o acumulación de dos tarjetas amarillas. Faltas son la cantidad de acciones que infringen el reglamento y el árbitro decide señalar. Tiros de esquina se tratan de las reanudaciones de juego concedidas al ataque cuando el balón sale por la línea de fondo y fue tocado por la defensa.

Centros es el número de pases largos o elevados enviados desde las bandas hacia el área rival. Toques de balón se refiere a la cantidad de veces que los jugadores de un equipo tocan el balón. Entradas representan las acciones defensivas para quitar el balón al oponente mediante una acción física. Intercepciones consiste en el número de veces que un jugador corta un pase del equipo contrario anticipándose a la jugada. Duelos aéreos ganados es la cantidad de disputas por el balón en el aire que un equipo gana. Despejes se tratan de las acciones defensivas para alejar el balón del área propia. Fueras de lugar representa la infracción en la que un jugador se encuentra en posición adelantada al momento de recibir el balón. Saques de meta es la reanudación del juego desde el área chica cuando el balón sale por la línea de fondo y fue tocado por el ataque. Saques de banda son las reanudaciones del juego desde las bandas laterales cuando el balón sale por los costados. Y balones largos representan los pases de larga distancia desde la defensa al ataque.

Es importante señalar que, el modelo en valores absolutos no tenía significancia alguna, es decir, ninguna de las variables en valor absoluto resultó ser significativa, por lo que se procedió a trabajar con las diferencias respecto al rival (valor de la variable de Tigres-valor de la variable del rival) en cada uno de los estadísticos, con excepción de la variable binaria de la localía (toma el valor de 1 cuando Tigres es local y 0 cuando es visitante). Lo cual captura de manera relativa el desempeño de un equipo de fútbol y puede impactar directamente en la probabilidad de victoria, por ejemplo, no es lo mismo realizar cuatro tiros a portería cuando el rival hizo dos, a realizar los mismos cuatro cuando el rival tuvo ocho.

No se omite mencionar que el modelo final no incluye todas las variables explicativas de la Figura 1, solamente incorpora las variables significativas, donde la selección de variables se basó en criterios teóricos y estadísticos para asegurar un modelo parsimonioso y evitar con ello un potencial sobreajuste.

Criterios de inclusión/exclusión

Se han incluido en el estudio todos los juegos correspondientes a jornada regular del torneo nacional, por lo que han sido excluidos aquellos partidos de fase final (liguilla), así como partidos de índole internacional y partidos amistosos de cualquier tipo.

Consideraciones éticas

Debido a que se utilizaron datos de acceso público, no se requirió de una aprobación ética dado que no involucra participantes humanos directamente ni datos sensibles.

Procedimiento

Inicialmente, se recopilaron los datos estadísticos de cada partido del equipo Tigres durante la temporada regular 2024-2025 desde Sports-Reference. Despues, se construyeron las variables del modelo, principalmente a partir de la diferencia estadística entre Tigres y su rival en cada partido, salvo para la variable binaria de localía. Luego, se aplicó una regresión logística binaria para estimar la probabilidad de victoria del equipo, usando como variable dependiente la victoria (1) o no victoria (0). Posteriormente, la estimación del modelo se llevó a cabo mediante el método de máxima verosimilitud. Todo el análisis fue realizado en el lenguaje de programación R (versión 4.2.3), utilizando funciones estándar para regresión logística y manejo de datos.

En particular, la estimación del modelo de regresión logística se efectuó utilizando la función `glm` del paquete base de R, complementada con los paquetes `margins` para el cálculo de efectos marginales y `pscl` para la bondad del ajuste. La evaluación de supuestos incluyó el análisis de multicolinealidad mediante `car` y `performance`,

Análisis estadístico de datos.

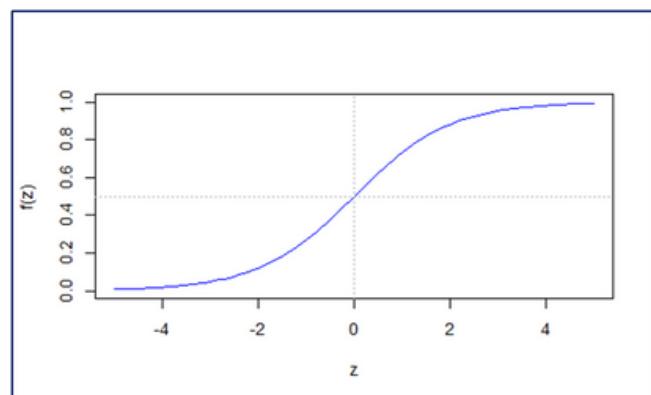
A continuación se plasma el marco metodológico en el cual se realiza el análisis estadístico. El modelo logit tiene como fundamento matemático a la función logística (Wooldridge, 2013). La cual puede expresarse como:

$$f(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}$$

Donde el argumento puede tomar los valores de z y la función se encuentra en un rango de 0 a 1 , lo cual es sumamente útil en probabilidad y estadística, esto puede observarse en la siguiente figura.

Figura 2

Gráfica de la función logística utilizada en el modelo



La Figura 2 representa la curva sigmoidal típica que se usa en regresión logística, la cual modela la distribución acumulada de una variable aleatoria logística estándar.

El modelo logit consiste en la estimación de la probabilidad de que ocurra un evento binario, (Gujarati y Porter, 2010).

Por lo tanto, puede establecerse que si es la probabilidad de que ocurra un evento dicotómico, en éste caso, la victoria del equipo, y es un modelo lineal, se puede representar dicha probabilidad como:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i X_i)}}$$

Donde es la probabilidad de que ocurra el evento , mientras que, representa a las variables independientes, es decir:

$$P_i = E[Y_i = 1 | X_i]$$

Como puede observarse, no existe una relación lineal entre y , además, no existe una linealidad en parámetros, lo cual no permite la estimación mediante mínimos cuadrados ordinarios (MCO). Por lo tanto, es necesario linealizar los parámetros.

Si es la probabilidad de obtener una victoria, la probabilidad de no alcanzar la victoria es . Por lo tanto, podemos expresar que el logaritmo natural de la razón o proporción de las probabilidades de obtener una victoria respecto a no obtenerla son:

$$LOGIT_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \sum_{i=1}^n \beta_i X_i$$

Donde representa la función a estimar (el modelo logit), el cual es lineal en parámetros y respecto a .

Una vez linealizados los parámetros, se proceden a estimar los coeficientes mediante el método de máxima verosimilitud, el cual estima los valores de los parámetros que maximizan la verosimilitud de la muestra observada (Wackerly et al., 2008). Dado que, la variable sigue una distribución binomial, la función de máxima verosimilitud es:

$$L(\beta) = \prod_{i=1}^n P_i^{Y_i} (1 - P_i)^{1 - Y_i}$$

Donde representa a los parámetros de la función, y el método tiene como objetivo maximizar el logaritmo natural de esta función de máxima verosimilitud.

Por su parte, la validación del modelo consiste en el cumplimiento de la no multicolinealidad, lo cual será detectado mediante el factor de inflación de la varianza (VIF, por sus siglas en inglés), donde se esperan valores entre 1 y 5. La autocorrelación será evaluada mediante un análisis gráfico de la función de autocorrelación de los residuos de Pearson, donde se espera que los valores sean menores a 0.33 dado el tamaño de muestra.

El ajuste del modelo será cuantificado a través de la prueba Box-Tidwell para analizar la linealidad, donde se espera un valor $p > 0.05$. Así como por el área bajo la curva (AUC) para examinar la especificación del modelo, donde se espera un valor cercano a 1. Y por la r-cuadrada de McFadden que indica la bondad del ajuste, donde se espera un valor cercano a 1. Lo anterior, contemplando la ausencia de valores atípicos influyentes, lo cual se puede constatar en la siguiente sección.

Por otro lado, la elección del modelo óptimo se realizó paso a paso (stepwise) mediante una eliminación hacia atrás (backward), de lo general a lo particular, en la primera etapa se estimó el modelo completo, en la siguiente se omitió la variable con el coeficiente menos significativo, y así sucesivamente hasta contar solamente con coeficientes significativos en el modelo.

Resultados

Antes de mostrar los resultados del análisis de regresión logística, se procede a mostrar la estadística descriptiva de las variables.

Tabla 1*Estadística descriptiva de las variables del estudio*

Variable	Tigres				Rival			
	Promedio	Desviación estándar	Valor mínimo	Valor máximo	Promedio	Desviación estándar	Valor mínimo	Valor máximo
Posesión de balón	54	12	30	79	46	12	21	70
Tarjetas amarillas	194	152	0	5	241	162	0	6
Tarjetas rojas	18	39	0	1	9	38	0	2
Faltas	12	4	7	20	13	3	5	21
Tiros de esquina	6	3	1	14	4	2	0	10
Centros	22	9	7	52	15	6	6	31
Toques de balón	621	119	379	941	556	116	329	770
Entradas	18	5	7	28	16	5	7	29
Intercepciones	10	5	4	22	9	4	2	16
Duelos aéreos	12	4	5	19	12	4	4	23
Despejes	18	6	6	29	28	10	8	49
Fueras de lugar	2	2	0	6	1	1	0	3
Saques de meta	8	4	2	19	10	4	3	23
Saques de banda	22	5	13	35	20	6	7	32
Balones largos	67	10	44	82	70	12	49	95

La Tabla 1 contiene la estadística descriptiva para cada una de las variables contenidas en el estudio, donde la variable localía es una variable dicotómica y en el periodo considerado, como se trata de un año completo, existen 17 partidos de local y 17 de visitante. Respecto a dicha estadística, se tiene que destacar que Tigres presentó en promedio, 8% más de posesión que el rival a lo largo de la temporada, menor cantidad de amarillas (0.47) pero mayor número de rojas (0.09), menor cantidad de faltas (1), despejes (10), saques de meta (2) y balones largos (3), pero mayor cantidad de tiros de esquina (2), centros (7), toques (65), entradas (2), intercepciones (1), duelos aéreos ganados (1), fúera de lugar (1) y saques de banda (2).

Tabla 2

Resultados del modelo

Variable	Coeficiente estimado	Error estándar	Intervalo de confianza	Efectos marginales	Factor de inflación de
Intercepto	-789.188	788.746	(-2.201517, 0.465236)	-	-
Local	4.659474***	1.794.805	(2.234592, 8.368087)	5.358	294
Toques	0.007999*	4.489	(0.001368, 0.016552)	9	311
Entradas	0.155347*	97.162	(0.008749, 0.339046)	178	174
Amarillas	0.778687**	396.312	(0.201258, 1.545945)	895	151
Centros	-0.201068**	88.316	(-0.372684, -0.074925)	-231	425
R-cuadrada de McFadden				4.728.844	

Nota. * indica que el coeficiente estimado es significativo al 10%, ** al 5% y *** al 1%.

Los resultados del modelo se plasman en la Tabla 2, los cuales indican que cinco variables tienen un efecto estadísticamente significativo en la probabilidad de victoria. La localía es el factor con el impacto positivo más fuerte, mientras que un mayor número de toques, entradas y tarjetas amarillas que el rival también incrementan la probabilidad de ganar. Por el contrario, generar más centros al área que el oponente se asocia con una disminución en la probabilidad de triunfo.

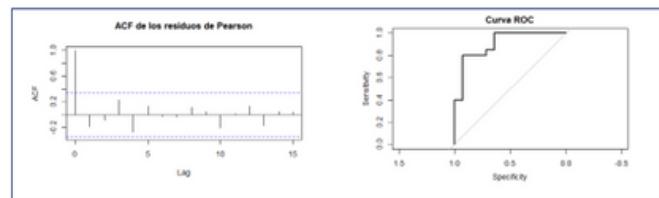
Es decir, la variable de mayor relevancia en el modelo es la cancha donde se disputa el partido, ya que los resultados indican que jugar de local incrementa en 53.58 puntos porcentuales la probabilidad de victoria, en promedio y manteniendo todo lo demás constante.

Mientras que, alcanzar un mayor número de toques de balón que el rival se asocia a un incremento en la probabilidad de ganar un partido, por lo que obtener 100 toques más que el rival se refleja en un incremento del 9.19 puntos porcentuales en la probabilidad de victoria. Esto puede interpretarse como la eficacia de la capacidad técnica de los jugadores, así como de la propuesta táctica del entrenador. Por otro lado, el número de entradas se asocia a un incremento en las probabilidades de alcanzar el triunfo en el partido, donde realizar 10 entradas más que el rival incrementa la probabilidad de ganar en 17.86 puntos porcentuales, lo cual puede representar el nivel de superioridad en intensidad del equipo respecto al rival. Relativo al número de tarjetas amarillas, por cada tarjeta adicional que el equipo recibe respecto al equipo contrario, la probabilidad de obtener una victoria se incrementa en 8.95 puntos porcentuales, lo cual puede reflejar el nivel de agresividad del equipo y los resultados positivos que ello implica. Finalmente, respecto al número de centros, puede notarse que un mayor número de centros que el rival se relaciona a una reducción de las probabilidades de ganar, por cada 10 centros más que el rival, las probabilidades disminuyen en 23.12 puntos porcentuales. Esto puede entenderse desde el punto de vista táctico, donde el juego en conjunto mediante paredes, triangulaciones, diagonales, pases filtrados y en general, las transiciones ofensivas interiores, representan un impacto positivo en el resultado final de un partido. Se ha utilizado el factor de inflación de la varianza (VIF) para detectar multicolinealidad entre las variables explicativas, donde se ha encontrado baja colinealidad para cada una de las variables. Además, no fue necesario ninguna transformación, debido a que todas las variables cumplen con la linealidad del modelo logit, de acuerdo a la prueba de Box-Tidwell ($p > 0.05$).

Así mismo, se ha probado la autocorrelación de los residuos y no se ha encontrado evidencia de ella mediante el gráfico de la función de autocorrelación, lo cual se muestra en la Figura 3. Por otro lado, se ha estimado la curva característica de operación del receptor (ROC, por sus siglas en inglés) para evaluar el desempeño del modelo, lo cual se muestra en la Figura 3, obteniendo un área bajo la curva (AUC) de 0.9036, lo cual indica que se trata de un modelo con muy buena capacidad de predicción. En este sentido, la r -cuadrada de McFadden calculada indica que se trata de un ajuste adecuado del modelo.

Figura 3

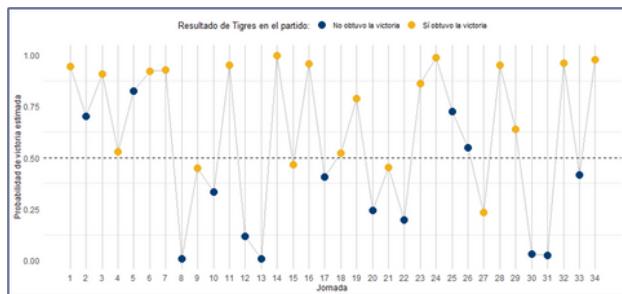
Análisis gráfico de validez del modelo



Para complementar los resultados de la Tabla 2 y con el fin de validar el modelo, se muestra la siguiente figura que contempla la probabilidad estimada de triunfo por partido.

Figura 4

Probabilidad estimada de victoria por partido



La Figura 4 ilustra las probabilidades estimadas de obtener una victoria mediante el modelo logit propuesto para el equipo de Tigres en la temporada 2024-2025, donde puede observarse un grado de precisión del 76.47%, considerando un umbral límite del 50%. De las 20 victorias predichas, solamente se cumplieron 16, mientras que, de los 14 partidos que se estimaron donde no habría una victoria, no se alcanzó el triunfo en 10. Es decir, en cuatro partidos se estimó incorrectamente la victoria e igualmente, en cuatro partidos se predijo incorrectamente la no victoria.

Figura 5

Matriz de confusión

		Real	
		Victoria	No victoria
Predicho	Victoria	16	4
	No victoria	4	10

En la Figura 5 se muestra la matriz de confusión, donde se plasman los resultados predichos y reales obtenidos por el equipo de Tigres, donde se resalta el 76.47% de precisión y una sensibilidad del 80%. Es importante destacar lo acontecido estadísticamente en dichos partidos donde no se cumplió lo predicho, lo cual se puede encontrar en la siguiente sección, en donde se discuten los resultados obtenidos.

Discusión

El principal resultado del estudio indica que las variables local, toques, entradas, amarillas y centros, fueron las que resultaron significativas, por lo cual estos resultados confirman lo encontrado por Martínez Martínez y González García (2019) y Arrunategui Zeta (2021), respecto a la ventaja de jugar un partido de local. A pesar de que dichos estudios han sido conducidos en distintos contextos a nivel competitivo, ya que el primero trabajó con las cinco principales ligas de Europa (Inglaterra, España, Italia, Alemania y Francia) con datos entre 2016 y 2017 y el segundo con la liga de Perú con datos de 2015 a 2019. Con esto se puede decir que la condición de local juega un papel determinante, independientemente del espacio-tiempo.

Así mismo, se confirma que la posesión de balón no incrementa las probabilidades de obtener una victoria, al igual que lo señalado por Botana García (2022), quien analizó los partidos entre 2016 y 2021 de las cinco principales ligas de Europa, con lo cual se puede inferir que la posesión no representa un elemento clave en la obtención de una victoria, quizás en el puntaje final del torneo o en la consecución de algún título o de alguna meta como la salvación por el descenso, pero eso sería digno de analizarse en otro estudio.

No obstante, el número de toques de balón resultó ser un determinante de la victoria en México, lo cual está relacionado con el control de juego y el dominio territorial que puede llegar a tener un equipo, pero dicho control tiene impacto si se trabaja con la circulación de balón, es decir, no se trata de solo tener el balón de forma parsimoniosa, sino de tenerlo y buscar distintas alternativas para desarrollar el modelo de juego.

Al igual que Lepschy et al. (2020), se ha determinado que los centros realizados por un equipo se relacionan negativamente con la probabilidad de obtener un triunfo, por lo que, se puede inferir que en México y en Alemania, reducir el número de centros incrementa la probabilidad de victoria. Esto debe estar vinculado con la construcción y el desarrollo del modelo de juego, ya que adoptar al centro como finalización recurrente de jugadas ofensivas ciertamente limita la creatividad de un equipo al ataque, lo cual facilita las labores defensivas del equipo rival. Por lo tanto, construir transiciones ofensivas que incluyan una amplia diversidad de opciones, impacta en el resultado obtenido. En otras palabras, atacar con paredes, triangulaciones, combinaciones, permutaciones, pases filtrados, diagonales retrasadas, diagonales adelantadas, entre otras variantes, repercuten en la probabilidad de victoria de un equipo.

Ahora bien, estos resultados contrastan con lo encontrado por Chocano Feito y Castilla González (2021), quienes determinaron que en las ligas de Alemania y España los tiros de esquina son una variable determinante para el resultado en un encuentro, lo cual puede estar asociado al alto nivel competitivo que existe en estos países, donde la precisión representa un factor clave.

Por su parte, en éste estudio no se determinó que las tarjetas rojas son un elemento determinante en la consecución de un triunfo, como si lo concluyeron Chocano Feito y Castilla González (2021), sin embargo, las tarjetas amarillas se asocian a la probabilidad de una victoria pero dicho efecto es positivo, es decir, a mayor número de tarjetas amarillas, mayor probabilidad de triunfo, con lo cual se puede inferir que las tarjetas amarillas son la variable observable pero detrás de ello se encuentra un proceso subyacente asociado a la intensidad defensiva con la que juega un equipo, ya que no es casualidad que el equipo al incrementar el número de entradas y de tarjetas amarillas, incrementa la probabilidad de obtener una victoria. Si bien, el modelo mostró un alto grado de precisión, cabe destacar los partidos en los cuales el modelo no fue acertado. En la jornada 2 del Apertura 2024, Tigres se enfrentó a Atlas y se estimó una probabilidad de 60.88% de obtener una victoria, sin embargo, dicho partido acabó en empate 1-1. En la jornada 5, la probabilidad de victoria fue del 93.04% contra el Guadalajara, solo que el encuentro concluyó con un empate 1-1. En la jornada 15, se enfrentó a Tijuana de visitante con un 41.11% de probabilidad de ganar, sin embargo, el partido acabó 0-3 a favor de Tigres. En la primera jornada del torneo Clausura 2025, Tigres visitó al Atlético San Luis y a pesar de tener el 44.22% de probabilidad de victoria, se llevó el triunfo 1-3. En la jornada 8 del Clausura 2025, la 25 del año, Tigres recibió a León, y tuvo una derrota 0-1 a pesar del 61.15% de probabilidad de victoria. En la jornada 10 del Clausura, 27 del año, Tigres visitó a Necaxa y con un 40.27% de probabilidad de victoria resultó ganador 1-2. En la jornada 29 del año, 12 del Clausura, Tigres recibió al Santos con una probabilidad de obtener el triunfo del 36.96% pero alcanzó la victoria 3-0.

Finalmente, en la penúltima jornada del año, Tigres recibió al Pachuca, y obtuvo el empate 0-0 a pesar de una probabilidad de victoria del 53.97%.

En éste sentido, el modelo tiene sus limitaciones importantes, donde el tamaño de muestra es la principal, lo cual puede repercutir en el poder predictivo del modelo, no obstante, se cumple con lo establecido empíricamente, contar con al menos 10 eventos por predictor. Dicha limitación obliga a adoptar el presente modelo como un modelo exploratorio en México. Por otro lado, la validez externa del modelo está limitada por el hecho de que los datos corresponden exclusivamente a la temporada 2024-2025 de Tigres. Por lo que, los resultados obtenidos son aplicables únicamente al contexto de dicha temporada y equipo. Para fortalecer la generalización de los hallazgos, sería necesario extender el análisis a múltiples temporadas o incorporar información de otros equipos de la liga. No obstante, el uso de variables de rendimiento relevantes y el sustento teórico proporcionado por la literatura especializada contribuyen a sentar bases para futuras investigaciones comparativas. Desde una perspectiva aplicada, los hallazgos del estudio ofrecen implicaciones directas para la gestión deportiva. La identificación de los factores que incrementan la probabilidad de victoria permite orientar decisiones estratégicas en áreas como la planificación del entrenamiento, la selección de jugadores, el diseño del modelo de juego y la definición de métricas internas de rendimiento. Para la dirección deportiva, estos resultados pueden servir como insumo en procesos de evaluación del plantel, scouting y toma de decisiones durante el mercado de transferencias. Para el cuerpo técnico, los indicadores asociados a la victoria pueden integrarse en la preparación táctica y psicológica del equipo, priorizando acciones que aumenten la probabilidad de éxito competitivo.

Es importante señalar que, al tratarse de un estudio observacional, los resultados del modelo reflejan asociaciones estadísticas entre los indicadores de rendimiento deportivo y la probabilidad de victoria del equipo. Por lo tanto, no es posible establecer relaciones de causalidad directa, sino únicamente correlaciones que podrían servir como base para futuras investigaciones experimentales o modelos más complejos que controlen variables omitidas. Incluso, no se puede descartar endogeneidad.

Conclusiones

El presente estudio tuvo como objetivo determinar la probabilidad de que el equipo de Tigres obtenga una victoria durante la temporada 2024-2025, por lo cual se utilizó un modelo logit con variables asociadas al rendimiento deportivo del equipo.

Como se ha señalado anteriormente, el principal hallazgo de la investigación radica en la determinación de los factores de rendimiento deportivo que se asocian a la probabilidad de victoria del equipo de Tigres, resultados que se pueden resumir de la siguiente manera.

La localía resultó ser la variable de mayor relevancia en el modelo. Mientras que, el número de toques de balón, número de entradas al rival y el número de tarjetas amarillas recibidas, todas medidas respecto a la diferencia de dicha estadística del equipo rival, tuvieron una relación positiva con la probabilidad de victoria del equipo. Siendo la variable del número de centros, la de asociación negativa a dicha probabilidad.

F La realización de este estudio aporta información empírica valiosa para los distintos tomadores de decisiones en el mundo del fútbol. Tiene aplicaciones prácticas para la mejora de toma de decisiones estratégicas en la gestión deportiva y en la planificación del entrenamiento, tanto para la parte física, como para la táctica e incluso la psicológica, ya que se genera estadística inferencial. Estos resultados permiten a directores deportivos, entrenadores y analistas optimizar procesos de scouting, valorar la eficacia de estilos de juego y ajustar intervenciones durante la temporada con base en evidencia estadística. Asimismo, la generación de estadística inferencial fortalece la construcción de un modelo de juego para el equipo de Tigres basado en toques de balón, con alto nivel de intensidad en la recuperación del esférico y con variantes ofensivas que reduzcan la dependencia del centro al área como principal recurso.

Como ya se ha comentado, la principal limitación del estudio ha sido el tamaño de muestra, lo cual incide en el poder predictivo del modelo. Por lo tanto, en futuros estudios se recomienda la incorporación de un mayor número de equipos, períodos y variables control, más allá de las asociadas al rendimiento deportivo, dichas incorporaciones repercutirían en la apropiada selección del modelo estadístico. Asimismo se sugiere la realización de investigaciones futuras en otras ligas con el fin de contrastar los hallazgos aquí obtenidos.

Referencias

- Arrunategui Zeta, M. A. (2021). Factores que influyen en la probabilidad de ganar un partido de fútbol: Análisis de la Liga 1 peruana en el periodo 2015–2019 [Tesis de maestría, Universidad Ricardo Palma]. <https://www.urp.edu.pe/pdf/id/34319/n/3-arrunategui-tesis-version-de-prepublicacion.pdf>
- Botana García, A. (2022). El impacto de la posesión en el fútbol moderno. [Trabajo de fin de máster, Universitat Pompeu Fabra–Barcelona School of Management]. <http://hdl.handle.net/10230/54809>
- Chocano Feito, P. J., & Castilla González, E. (2021). Estadística multivariante aplicada al análisis y predicción de partidos de fútbol en las principales ligas europeas. *Pensamiento Matemático*, 11(2), 21-30.
- Elyakim, E., Morgulev, E., Lidor, R., Meckel, Y., Arnon, M., & Ben-Sira, D. (2020). Comparative analysis of game parameters between Italian league and Israeli league football matches. *International Journal of Performance Analysis in Sport*, 20(2), 165–179. <https://doi.org/10.1080/24748668.2020.1726158>
- Fernández-Cortés, J. A., Cáceres Sánchez, L., Antunez, A., García-Rubio, J., & Ibáñez, S. J. (2022). Análisis de la influencia de las variables situacionales en el fútbol profesional. *Retos*, 46, 114–119. <https://doi.org/10.47197/retos.v46.91541>
- Gujarati, D., & Porter, D. (2010). *Econometría* (5.ª ed.). McGraw-Hill Educación.
- Lepschy, H., Wäsche, H., & Woll, A. (2020). Success factors in football: an analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, 20(2), 150–164. <https://doi.org/10.1080/24748668.2020.1726157>
- Marques Junior, N. K. (2024). Matveev's sportive periodization: Microcycle concept. *Marathon*, 16(2), 115–136. <https://doi.org/10.24818/mrt.24.16.02.04>

Martínez Martínez, F. D., & González García, H. (2019). Efecto de marcar primero y la localización del partido en las principales ligas del fútbol europeo. *Retos*, 35, 242-245. <https://doi.org/10.47197/retos.v0i35.64026>
Matveyev, L. P. (1977). Periodización del entrenamiento deportivo. Instituto Nacional de Educación Física.

Ovando Fuentealba, R. (2025). Predicción de resultados de partidos de la liga profesional de fútbol chileno usando algoritmos de machine learning [Tesis de maestría, Universidad del Desarrollo]. <https://repositorio.udd.cl/items/934ac3fe-4953-4691-a3db-c803ab006367>
Sports-Reference. (2025). 2024-2025 UANL Stats.

<https://fbref.com/en/squads/d9e1bd51/2024-2025/UANL-Stats>

Tamarit, X. (2007) ¿Qué es la periodización táctica? MC Sports.

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). Estadística matemática con aplicaciones (7.^a ed.). Thomson Learning.

Wooldridge, J. M. (2013). Introducción a la econometría: Un enfoque moderno (5.^a ed.). South-Western Cengage Learning.

12