



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

DETERMINAR EL DESEMPEÑO DE JUGADORES DE FÚTBOL MEDIANTE
APRENDIZAJE AUTOMÁTICO: UNA REVISIÓN SISTEMÁTICA

PROPUESTA DE TEMA DE MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN COMPUTACIÓN

CRISTIAN LILLO CIERO

MODALIDAD:
DOBLE TITULACIÓN CON MAGÍSTER

PROFESOR GUÍA:
JAVIER BUSTOS

SANTIAGO DE CHILE
2024

1. Introducción

Contexto y Motivación

Debido a la modernización del fútbol profesional, que en los últimos 15 años ha incorporado el uso del sistema de detección automática de goles (DAG) en las porterías (FIFA, 2012) y el videoarbitraje (VAR) en los partidos (International Football Association Board, 2018), ha existido un aumento en el uso y desarrollo de tecnologías que permitan recopilar automáticamente las acciones realizadas por los jugadores durante los partidos (De Sousa et al., 2011; Harell & Bajic, 2020). Esta información permite analizar el rendimiento tanto de los equipos como de los jugadores, y es utilizada por los clubes de fútbol para mejorar su desempeño y tomar decisiones informadas respecto a la contratación de jugadores (Wakelam et al., 2022).

Previo a este avance tecnológico, los únicos datos disponibles eran los que se encontraban en las planillas de los partidos, donde se registraban los goles, tarjetas y cambios de jugadores (Muñoz et al., 2024). Sin embargo, en la actualidad, existen distintos métodos para la recopilación de datos, donde destacan los sistemas de seguimiento óptico y los eventos de fútbol. Los sistemas de seguimiento óptico utilizan cámaras de alta velocidad para capturar la posición de los jugadores y del balón en el campo de juego, lo que permite obtener información detallada sobre la velocidad, aceleración y distancia recorrida por los futbolistas. Por otro lado, los eventos de fútbol recopilan información sobre las acciones realizadas con el balón, tales como pases, tiros, goles, faltas, entre otros. Cada uno de estos eventos es registrado incluyendo el momento en que ocurrió, la ubicación en el campo de juego, el tipo de acción y el jugador involucrado, donde además se pueden agregar propiedades adicionales dependiendo del tipo de acción, como la ubicación final de un pase o el resultado de un tiro (Decroos, 2020). La investigación de esta tesis se enfocará en los datos obtenidos a través de los eventos de fútbol, ya que en estos se entrega una descripción más detallada sobre lo ocurrido en los partidos y es además la más utilizada en la actualidad debido al menor costo de implementación (Decroos et al., 2019). Esta información suele ser provista por empresas especializadas en la recopilación de datos de fútbol, como Opta Sports, Wyscout y StatsBomb, y es utilizada por la mayoría de los modelos de Machine Learning y Deep Learning que buscan determinar el rendimiento de los jugadores (Pappalardo et al., 2019).

Problema

Para determinar el rendimiento de un jugador es necesario enfocarse en los eventos del partido donde participó el futbolista en particular. Pero, como no todas las acciones tienen la misma importancia en el resultado final del partido, es necesario determinar cuáles son las acciones clave a utilizar para evaluar el desempeño de un jugador, ya que no existe

un estándar objetivo al respecto (Pappalardo, Cintia, Ferragina, Massucco, Pedreschi, & Giannotti, 2019). Algunos autores se han enfocado en una sola acción, como los pases realizados en una jugada que termina en un tiro (Duch et al., 2010) o la importancia que tiene un pase de terminar en gol (Brooks et al., 2016), mientras que otros han considerado todos los eventos posibles y les han asignado un peso para determinar su influencia en el rendimiento final del jugador (Pappalardo, Cintia, Ferragina, Massucco, Pedreschi, & Giannotti, 2019). Es por esto que se hace necesario realizar una revisión sistemática a los artículos de investigación publicados en el área, a fin de definir variables generales y comparar el rendimiento de diferentes modelos de Machine Learning y Deep Learning.

2. Estado del Arte

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua quaerat. Ni idea. (Brinkjans et al., 2024)

Este es un ejemplo de una revisión sistemática en el fútbol, pero relacionada al valor monetario de los jugadores. Específicamente al costo de transferencia entre clubes (Franceschi et al., 2024)

Se muestra una clasificación de rendimiento mediante el uso de Plus-Minus, que es también utilizado en deportes como volleyball y hockey sobre hielo. (Kharrat et al., 2020)

3. Preguntas de Investigación

- ¿Se obtienen resultados más similares a los de Opta Sport y Wyscout con modelos de Machine Learning o de Deep Learning?
- ¿Es mejor utilizar variables obtenidas directamente de los eventos de fútbol o es preferible utilizar variables generadas a partir de estas?
- ¿Cuál es el mejor método para evaluar jugadores? ¿Considerando todas sus acciones o aquellas relevantes para su rol en el campo de juego?

4. Hipótesis

Es posible determinar variables, basadas en las acciones de jugadores en partidos de fútbol, que influyen de forma transversal en la evaluación de un futbolista profesional según su rol en el campo de juego.

5. Objetivos

Objetivo General

El objetivo de esta investigación es analizar las distintas estrategias utilizadas por investigadores del área para definir las acciones clave que sirven para determinar el rendimiento de un futbolista profesional en uno o varios partidos.

Objetivos Específicos

1. Recopilar artículos de investigación que utilicen datos de eventos de fútbol para determinar el rendimiento de los jugadores.
2. Identificar las variables utilizadas por los autores para evaluar el desempeño deportivo de los futbolistas.
3. Definir grupos de variables que sean comunes en los distintos artículos seleccionados.
4. Analizar influencia de las variables en cada uno de los artículos seleccionados.
5. Determinar variables generales, llamadas acciones clave, que permiten calificar el rendimiento de un jugador.

6. Metodología

La metodología de esta tesis puede ser dividida en las siguientes etapas:

Búsqueda de Artículos

Se realizará una revisión sistemática de la literatura, donde se buscarán artículos que utilicen datos de eventos de fútbol para determinar el rendimiento de los jugadores. Se utilizarán las bases de datos Scopus, IEEE Xplore, Web of Science y Google Scholar para la búsqueda, donde se filtrarán los papers mediante las palabras clave «soccer», «football», «player performance», «machine learning» y «deep learning».

Se seleccionarán los artículos que cumplan con los siguientes criterios de inclusión: (1) que utilicen datos de eventos de fútbol, (2) que determinen el rendimiento de los jugadores, (3) que utilicen modelos de Machine Learning o Deep Learning, (4) que estén escritos en inglés o español y (5) que estén publicados en los últimos 15 años. Se excluirán de la selección aquellos que no cumplan con los criterios también aquellos que no estén disponibles en texto completo.

Análisis de Artículos

Se extraerán las variables utilizadas por los autores para determinar el desempeño deportivo de los jugadores y se agruparán en categorías comunes. Posteriormente, se analizará la influencia de las variables en cada uno de los artículos seleccionados, a fin de determinar cuáles son las más relevantes y también para detectar posibles patrones en la combinación de variables.

Experimentación

Se le entregarán distintos conjuntos de datos a modelos de Machine Learning y Deep Learning, propuestos previamente en la literatura, para determinar variaciones en la evaluación de los jugadores. En base a los resultados, se definirán las acciones clave para calificar el rendimiento de un futbolista.

7. Aporte de la Tesis

1. Estudio comparativo de las distintas estrategias propuestas por investigadores para determinar el rendimiento de un jugador de fútbol.
2. Definición de acciones clave para medir el desempeño de un futbolista.

Referencias

- Brinkjans, D., Paul, Y., Perl, J., & Memmert, D. (2024). The Success-Score in Professional Football: a metric of playing style or a metric of match outcome?. *International Journal of Computer Science in Sport*, 23(1), 54-79. <https://doi.org/10.2478/ijcss-2024-0004>
- Brooks, J., Kerr, M., & Guttag, J. (2016). Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49-55. <https://doi.org/10.1145/2939672.2939695>
- De Sousa, S. F., De A. Araújo, A., & Menotti, D. (2011). An overview of automatic event detection in soccer matches. *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, 0, 31-38. <https://doi.org/10.1109/WACV.2011.5711480>
- Decroos, T. (2020). *Soccer analytics meets artificial intelligence: learning value and style from soccer event stream data*.
- Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals: Valuing Player Actions in Soccer. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1851-1861. <https://doi.org/10.1145/3292500.3330758>

- Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the Performance of Individual Players in a Team Activity. *PLOS ONE*, 5(6), 1-7. <https://doi.org/10.1371/journal.pone.0010937>
- FIFA. (2012). *Circular no. 1315* (Número 1315). <https://digitalhub.fifa.com/m/616d0cf17182e0fd/original/mkuwd3qgdze4wxzqglsv-pdf.pdf>
- Franceschi, M., Brocard, J.-F., Follert, F., & Gouguet, J.-J. (2024). Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys*, 38(3), 577-600. <https://doi.org/https://doi.org/10.1111/joes.12552>
- Harell, A., & Bajic, I. V. (2020). The Data Gap in Sports Analytics and How to Close It. *Artificial Intelligence in Team Sports Workshop at The Thirty Fourth AAAI Conference on Artificial Intelligence*. <https://ai-teamsports.weebly.com/uploads/1/2/7/0/127046800/paper9.pdf>
- International Football Association Board. (2018). *Circular n° 12* (Número 12). https://downloads.theifab.com/downloads/circular-12_es?l=es
- Kharrat, T., McHale, I. G., & Peña, J. L. (2020). Plus-minus player ratings for soccer. *European Journal of Operational Research*, 283(2), 726-736. <https://doi.org/https://doi.org/10.1016/j.ejor.2019.11.026>
- Muñoz, O., Monroy, R., Cañete-Sifuentes, L., & Ramirez-Marquez, J. E. (2024). Automated Discovery of Successful Strategies in Association Football. *Applied Sciences*, 14(4). <https://doi.org/10.3390/app14041403>
- Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach. *ACM Trans. Intell. Syst. Technol.*, 10(5). <https://doi.org/10.1145/3343172>
- Wakelam, E., Steuber, V., & Wakelam, J. (2022). The collection, analysis and exploitation of footballer attributes: A systematic review. *Journal of Sports Analytics*, 8(1), 31-67. <https://doi.org/10.3233/JSA-200554>