



JUGADORES DE FÚTBOL SIMILARES

ESPECIALIZACIÓN EN CIENCIA DE DATOS

AUTOR: FEDERICO NAHUEL CIRIGLIANO

TUTOR: Lic. DIEGO ARIEL AIZEMBERG

12-11-2020

BUENOS AIRES - ARGENTINA

CONTENIDO

| | | |
|-----|-------------------------------------|----|
| 1. | INTRODUCCIÓN | 3 |
| 2. | ESTADO DEL ARTE | 4 |
| 3. | DEFINICIÓN DEL PROBLEMA | 5 |
| 4. | JUSTIFICACIÓN DEL ESTUDIO | 6 |
| 5. | ALCANCES DEL TRABAJO Y LIMITACIONES | 7 |
| 6. | HIPÓTESIS | 8 |
| 7. | OBJETIVO GENERAL | 9 |
| 7.1 | Objetivos Específicos | 9 |
| 8. | METODOLOGÍA A UTILIZAR | 11 |
| 8.1 | Técnicas | 11 |
| 8.2 | Herramientas | 11 |
| 9. | RESULTADOS | 12 |
| 9.1 | Fase I | 12 |
| 9.2 | Fase II | 14 |
| 9.3 | Fase III | 14 |
| 9.4 | Fase IV | 14 |
| 9.5 | Fase V | 15 |
| 9.6 | Fase VI | 16 |
| 10. | DISCUSIÓN | 18 |
| 11. | CONCLUSIONES | 19 |
| 12. | REFERENCIAS BIBLIOGRAFÍA | 20 |

1. INTRODUCCIÓN

El presente trabajo de investigación tiene por objetivo acercar una nueva solución dirigida al mundo del fútbol que logra detectar futbolistas similares a uno dado. Si bien la propuesta es simple, lograr resultados realmente útiles ante escenarios tan diversos entre sí es todo un desafío.

A su vez, sus beneficios son altamente rentables. En el ambiente del fútbol las transferencias de jugadores dominan cifras desorbitantes. Contar con una ventaja competitiva en este sentido no es un tema menor. De esta forma, un club económicamente modesto podría comprar un jugador por un cifra razonable y, si a través del tiempo el futbolista cumple con la proyección esperada, las ganancias podrían multiplicarse en forma abrupta. En resumen, la herramienta que se presenta a lo largo de este trabajo facilita al departamento de reclutamiento de los clubes a descubrir jugadores de cualquier liga del mundo con características futbolísticas similares a cualquier jugador consagrado en las mejores ligas del mundo.

Otro caso de uso podría ser cuando un club necesita reemplazar a una de sus piezas más importantes. Ya sea por motivos de edad, venta, lesión, sanción, etc. Con esta solución se podrían encontrar jugadores que poseen características futbolísticas similares al que se busca reemplazar. En estos casos el ojo no está puesto sobre el aspecto económico, sino que la búsqueda principal es reemplazar a un jugador por otro que pueda lograr su mismo (o mayor) nivel futbolístico. Dada la variabilidad de escenarios donde la herramienta puede ser de suma utilidad, se la considerara como una clara ventaja competitiva para el área de los clubes encargada de reclutar o descubrir nuevos talentos.

Actualmente, existen múltiples soluciones de este tipo. Algunas de ellas utilizan fórmulas matemáticas de alta complejidad, otras contemplan estimaciones futuras sobre el rendimiento de los futbolistas, etc. Pero ninguna de ellas es considerada como indiscutida. Es precisamente aquí donde se observa un vacío de solución y, por consiguiente, una gran oportunidad para ofrecer una solución robusta, contundente y versátil ante las distintas necesidades que puedan surgir.

Particularmente la solución que se va a tratar en el presente trabajo de investigación pretende contemplar los aspectos futbolísticos más determinantes para un jugador, valorarlos en función del rol que ocupa cada uno de ellos en el campo de juego y finalmente utilizar modelos estadísticos para lograr resultados altamente precisos.

Palabras Claves: futbolistas, similitud, aspectos futbolísticos.

2. ESTADO DEL ARTE

Existen múltiples criterios para determinar si un jugador es futbolísticamente similar a otro. La importancia del análisis radica básicamente en el ahorro de tiempo y dinero a la hora de reclutar futbolistas. Dos factores fundamentales para un trabajo de investigación. Ya no se deberán observar cientos de jugadores de fútbol para encontrar el reemplazante adecuado de la figura del equipo. Alcanzará con detectar a sus X (equis) futbolistas más similares. De esta forma el trabajo de seguimiento se focalizará en una lista bastante más acotada.

La disponibilidad de un *dataset* adecuado para abordar este tipo de investigación es un aspecto fundamental para el éxito del trabajo. Dentro del ámbito del fútbol, abundan datos relacionados a los eventos más relevantes de un partido como son los goles y las asistencias. Sin embargo, en los últimos 5 años se ha diversificado la recolección y el análisis de datos. Ya no sólo se consideran aquellos eventos relacionados directamente a oportunidades de gol, sino que además se analizan el resto de los eventos (pases, tarjetas amarillas, *fouls*, *corners*, etc) que suceden durante un partido de fútbol. A partir de este cambio de paradigma, se abren nuevas puertas para realizar diversas investigaciones. Por ejemplo, se puede determinar cuán influyente es cada pase dado en función a la probabilidad de gol que tiene asociada [1]. Otra novedad es la incorporación de datos asociados a características y decisiones de juego de cada jugador, lo cual permite determinar con mayor precisión el nivel de similitud entre dos futbolistas [2]. Esta es la temática que abordará el presente trabajo de investigación.

Actualmente, este tipo de soluciones son implementadas considerando a aquellos futbolistas que participan en las 5 ligas más reconocidas del mundo: *Premier League* (Inglaterra), *La Liga* (España), *Calcio* (Italia), *Bundesliga* (Alemania) y *Liga1* (Francia). El motivo es que las empresas que se ocupan de la recolección de datos observan únicamente las ligas más populares.

Distintos trabajos científicos coinciden en abordar la problemática de detectar jugadores de fútbol similares creando un vector multidimensional donde cada dimensión es el valor asociado a un aspecto futbolísticos del jugador. La elección de qué características de juego se considerarán depende fundamentalmente del responsable de investigación.

Otra decisión relevante es cómo calcular el grado de similitud entre dos futbolistas a partir de los vectores mencionados anteriormente. En algunos trabajos alcanza con calcular la distancia Euclidiana entre vectores [3]. Cuanto menor sea esta distancia, mayor será el grado de similitud. En cambio, existen otras investigaciones donde se crean fórmulas complejas cuyos resultados fueron estandarizados entre 0 y 1. Incluso, para comparar futbolistas de diferente edad, existen casos donde se realizan proyecciones a futuro en función de sus rendimientos actuales [4].

Otro criterio utilizado para calcular el nivel de similitud entre futbolistas, es determinar los patrones de circuitos de pases en los que un jugador ha logrado más participación. Basados en este concepto, se realizan comparaciones considerando el estilo de juego de los futbolistas [5].

Por último se pueden analizar los movimientos de los futbolistas dentro del campo de juego [6]. Para ello existen métodos automáticos, como cámaras especializadas, que capturan coordenadas recorridas por los jugadores a lo largo del partido. De esta forma se establecen las comparaciones considerando las áreas del campo en las que más participación tuvo un futbolista.

El procedimiento para determinar jugadores de fútbol similares es variado y depende fundamentalmente de cuáles son los aspectos futbolísticos considerados. Sería óptimo comprender la mayor cantidad de variables posibles para lograr un algoritmo que pueda aportar resultados sólidos y sea utilizado en el mundo real. La oportunidad es que actualmente no existe una forma aceptada e indiscutida para obtener este tipo de resultados.

3. DEFINICIÓN DEL PROBLEMA

No existe un único algoritmo para determinar el nivel de similitud entre jugadores de fútbol. Cada investigación utiliza su propio criterio para descubrir qué jugadores son futbolísticamente similares. En general, los resultados obtenidos difieren en función del algoritmo utilizado. Aún existe la posibilidad de descubrir un modelo que brinde resultados precisos ante escenarios diferentes.

No es lo mismo determinar si un arquero es futbolísticamente similar a otro que llevar a cabo dicho análisis para un delantero o un defensor. Todos poseen características de juego completamente distintas. Por otro lado, las edades de los futbolistas también son un factor importante a tener en cuenta. Quizás en la actualidad dos profesionales no sean similares entre sí, pero estableciendo una proyección a futuro se podría determinar si serán futbolísticamente similares.

El desafío es desarrollar un modelo exitoso considerando la mayor cantidad de escenarios posibles. De esta forma se podrán obtener resultados internacionalmente aceptados, ya que en la actualidad no existe una única solución para abordar dicha problemática.

4. JUSTIFICACIÓN DEL ESTUDIO

¿Cuánto tiempo y dinero podrían ahorrar los clubes si lograran determinar rápidamente qué futbolistas son los más adecuados para reemplazar a sus principales figuras? ¿Cuánto podrían disminuirse las diferencias jerárquicas si los clubes menos poderosos lograran conocer qué jugadores que no requieren contratos excéntricos son los más similares a determinados futbolistas de élite? Deberían hacerse preguntas como estas cuando se piensa en beneficios a la hora de implementar soluciones que puedan detectar jugadores de fútbol similares.

Sin dudas el departamento de reclutamiento de jugadores de un club es el área que más interactúa en este tipo de cuestiones. Una de las problemáticas que se pretende resolver desarrollando algoritmos que detectan jugadores de fútbol similares es evitar el procesamiento manual de grandes volúmenes de datos y videos para encontrar qué futbolistas son los adecuados para reemplazar a otros. Además, con este tipo de técnicas se logra ampliar considerablemente el espectro de futbolistas a analizar. En una sola ejecución del algoritmo, se podrán tener en cuenta la mayoría de los jugadores que participan en las principales ligas del mundo.

El objetivo de este tipo de soluciones es lograr resultados que de otra forma no podrían ser obtenidos. Analizar características de jugadores que participan en las ligas profesionales del mundo podría costar meses o años de trabajo. Por el contrario, los mercados de pases necesitan respuestas cada vez más precisas e inmediatas. La competencia entre los clubes del mundo por descubrir nuevos talentos es cada vez más intensa, ya que esto significa potenciales negocios multimillonarios. Contar con una ventaja competitiva en este sentido podría representar grandes beneficios.

5. ALCANCES DEL TRABAJO Y LIMITACIONES

El presente trabajo de investigación será llevado a cabo utilizando como base de datos la registrada para la última edición del videojuego *FIFA 19*. Dicha base de datos valora atributos futbolísticos de jugadores de todas las ligas profesionales del mundo. Es decir, contempla más de 700 clubes licenciados y supera los 17.000 jugadores de fútbol.

EA SPORTS (en adelante *EA*), empresa propietaria de *FIFA 19*, cuenta con un equipo de 6.000 ojeadores repartidos por todo el mundo. Estos expertos, una vez finalizado cada partido, acceden a la red interna de *EA* para actualizar una gran base de datos dinámica.

Una vez computados los datos entran en juego los 300 editores de *EA*. Los mismos asimilan y organizan el trabajo de los 6.000 ojeadores en campos de datos diferentes y 35 categorías de atributos específicos. Finalmente, estos datos derivan en una calificación global para cada jugador.

Cada jugador tiene una calificación global y seis puntuaciones en subcategorías claves: ritmo, tiro, pase, regate, defensa y físico. Dichas estadísticas se combinan con el reconocimiento internacional de un jugador para determinar la calificación global del jugador. Así se asegura una valoración acorde al nivel de exigencia profesional en el que se encuentra cada futbolista.

La puntuación global de cada jugador también se determina en función de su posición en el juego. Se ponderan sus atributos en función de la posición que ocupe dentro del campo de juego. Por ejemplo el remate al arco no se valora de la misma manera para un delantero que para un defensor.

Por último, cabe aclarar que los revisores y editores de datos de *EA* se encuentran permanentemente haciendo modificaciones en la base de datos del juego. Generalmente se tratan de cambios menores como modificaciones físicas (peinado, altura, peso, etc.), pero también puede suceder que se modifiquen estadísticas de un jugador durante la misma temporada.

6. HIPÓTESIS

Es posible detectar jugadores de fútbol similares entre sí respecto a sus características de juego. Partiendo de una base de datos donde se valoren múltiples características de juego, es factible generar fórmulas matemáticas que determinen quiénes son futbolísticamente similares.

Calculando el nivel de similitud entre dos futbolistas y estableciendo un umbral de decisión, se podrá considerar si los profesionales son similares entre sí. De esta forma, dado un jugador de fútbol se podrán detectar sus tres colegas futbolísticamente más similares.

Es importante que la base de datos mencionada anteriormente posea la mayor cantidad de métricas posible. Algunos ejemplos podrían ser posición en el campo de juego, edad del futbolista, remates al arco, agresividad, quite, velocidad, precisión, etc. Cuanto más métricas sean consideradas, más preciso será el nivel de similitud.

7. OBJETIVO GENERAL

El objetivo de la presente investigación es desarrollar un algoritmo que compare y valore distintas características futbolísticas de jugadores de fútbol. De esta forma se podrá determinar qué jugadores son futbolísticamente similares entre sí.

7.1 Objetivos Específicos

La investigación se llevará a cabo implementando distintas fases. Cada una de ellas finalizará al cumplir un objetivo específico. La sumatoria de los objetivos específicos significa el logro del objetivo general.

- ⇒ **Fase I:** Se formará una base de datos donde serán valorados los aspectos futbolísticos de jugadores profesionales. Es importante contemplar futbolistas de diferentes ligas del mundo para enriquecer el resultado de la investigación. Cuanto mayor sea la cantidad de futbolistas y características de juego contempladas, mayor será la precisión del algoritmo para determinar el grado de similitud entre futbolistas y también será mayor el espectro de jugadores involucrados a lo largo de la investigación.

Objetivo específico → Crear una base de datos donde se valoren aspectos futbolísticos de cada jugador.

- ⇒ **Fase II:** Con el objetivo de reducir la cantidad de variables a considerar dentro del cálculo matemático, se aplicará el concepto de componentes principales sobre los atributos de cada jugador. De esta forma se quitará la redundancia de información existente entre los datos logrando un nuevo conjunto de atributos no correlacionados entre sí.

Objetivo específico → Aplicar componentes principales a los atributos existentes en la base de datos.

- ⇒ **Fase III:** Se formará un vector en función de los valores resultantes de aplicar componentes principales a los atributos originales de cada futbolista. Es necesario que dichos valores originales de cada jugador sean expresados en la misma unidad de medida.

Objetivo específico → Formar vectores que representen matemáticamente las componentes principales obtenidas para cada futbolista.

- ⇒ **Fase IV:** Se desarrollará un cálculo matemático donde, a partir de distancias entre vectores y ponderación de atributos según posición ocupada dentro del campo de juego, se determine el nivel de similitud entre dos futbolistas. Además se deberá establecer un umbral donde a partir de dicho valor se podrá considerar que dos jugadores son futbolísticamente similares.

Objetivo específico → Desarrollar un cálculo matemático cuyo resultado sea el nivel de similitud entre dos futbolistas.

- ⇒ **Fase V:** A partir de ejemplos reales analizar los resultados obtenidos. Es decir, dado un jugador de fútbol, se detectarán los 3 jugadores futbolísticamente más similares utilizando el algoritmo desarrollado anteriormente.

TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

Objetivo específico → Ejecutar el algoritmo desarrollado utilizando casos reales.

⇒ Fase VI: Se compararán los resultados obtenidos con las características reales de los futbolistas.

Objetivo específico → Validar resultados obtenidos.

8. METODOLOGÍA A UTILIZAR

A continuación se describen las principales técnicas y herramientas utilizadas para lograr el desarrollo de la solución deseada.

8.1 Técnicas

Para lograr la implementación de un algoritmo que determine si dos jugadores son futbolísticamente similares entre sí, fue necesario recurrir a técnicas de diferentes niveles de complejidad que han sido útiles facilitando el recorrido hacia los resultados obtenidos.

En la etapa de pruebas de las distintas versiones del algoritmo planteado, fue madurando la idea de otorgar mayor peso a determinados atributos futbolísticos en función de la posición que ocupa un jugador dentro del campo de juego. Con esta modificación se logró obtener mayor precisión en los resultados, ya que a la hora de comparar futbolistas *se puso una lupa* en las características más importantes de un jugador según la posición que ocupa. Por ejemplo, en el caso de un defensor adquirieron mayor importancia los atributos de juego relacionados al quite del balón, intercepciones, marca, etc.

Otro recurso técnico que utiliza el algoritmo presentado es el Análisis de Componentes Principales (en adelante PCA). Debido a que la base de datos está conformada por 40 atributos futbolísticos, se planteó la idea de reducir la cantidad de variables eliminando la correlación que existe entre ellas. Observando la relación entre porcentaje de variabilidad de información conservada y cantidad de atributos, se decidió preservar 31 de los 40 atributos conservando de esta forma el 97% de la variabilidad de información. Otra decisión fue aplicar PCA utilizando la matriz de covarianza. En este último punto es donde se unen las dos técnicas aplicadas. Dado que se habían ponderado atributos otorgando mayor variabilidad, se optó por utilizar la matriz de covarianza a la hora de aplicar PCA para justamente resaltar en la consideración aquellos atributos que presentan mayor variabilidad de datos.

Finalmente con los valores resultantes de aplicar PCA ponderando determinados atributos futbolísticos, se conformaron los vectores multidimensionales que representaron a cada jugador. Dado que los atributos presentan la misma unidad de medida, se optó por determinar el nivel de similitud entre dos futbolistas calculando la distancia euclidiana entre los vectores que los representan.

8.2 Herramientas

Las herramientas utilizadas para lograr la solución presentada fueron: *PostgreSQL*, *R* y *Google Colab*. Todas ellas poseen la particularidad de ser gratuitas. En *Postgres* se montó la base de datos. Es decir el *input* principal del algoritmo, ya que posee las valoraciones de cada atributo futbolístico y no futbolístico de los más de 17.000 futbolistas que luego se compararán entre sí.

Se utilizó *R* como lenguaje de programación para desarrollar el algoritmo que determina el nivel de similitud entre los futbolistas. Además de tratarse de *software* libre, cuenta con librerías adecuadas para poder aplicar las técnicas descritas anteriormente.

Finalmente para presentar el trabajo realizado en forma unificada y estéticamente más amigable, se migró toda la lógica a un *notebook* de *Google Colab* utilizando *R* como lenguaje de programación.

En resumen, los datos de todos los futbolistas fueron almacenados inicialmente en una tabla de *Postgres*. Luego, aplicando modificaciones para que todos los atributos tengan la misma unidad de medida, se exportaron a un archivo *.csv*. Posteriormente se utilizó *R* para leer estos datos, transformarlos y crear la función que recibe como parámetro un jugador de fútbol y retorna sus 3

colegas futbolísticamente más similares. Dentro de esta función se encuentra la lógica propia de la solución propuesta durante el presente trabajo de investigación. Por último se migró toda la lógica a un *notebook* de *Google Colab* unificando los pasos mencionados anteriormente.

9. RESULTADOS

En esta sección se presentarán los resultados obtenidos durante las distintas fases definidas como objetivos de la investigación.

9.1 Fase I

Objetivo: Crear una base de datos donde se valoren aspectos futbolísticos de cada jugador.

Como resultado de la Fase I, a continuación se presenta la descripción del *dataset* utilizado para la investigación. Se trata de la base de datos del prestigioso juego de consola *FIFA 2019*, la cual refiere a valoraciones de distintos atributos futbolísticos de los jugadores profesionales [7]. De manera arbitraria, se han quitado aquellos campos que no se consideran útiles para la presente investigación. Por lo tanto, está compuesta por 18.207 registros con los siguientes 40 atributos:

| Atributos | Descripción |
|--------------|---|
| Overall | Puntaje General |
| Weak Foot | Pierna inhábil |
| Skill Moves | Habilidad |
| Work Rate | Esfuerzo |
| Position | Posición |
| Crossing | Habilidad para lanzar centros al área |
| Finishing | Precisión de tiro dentro del área |
| Heading | Juego aéreo |
| Accuracy | Precisión |
| ShortPassing | Precisión en pases cortos |
| Volleys | Precisión de voleas |
| Dribbling | Cuán bien se mueve con el balón |
| Curve | Cuánta comba puede lograr en tiros o pases |
| FKAccuracy | Precisión en tiros libres |
| LongPassing | Precisión en pases aéreos |
| BallControl | Control en el primer toque |
| Acceleration | Cuán rápido llega a su velocidad máxima |
| SprintSpeed | Velocidad máxima |
| Agility | Velocidad del control del balón al correr o al cambiar de dirección |
| Reactions | Velocidad de reacción |
| Balance | Equilibrio |
| ShotPower | Cuán fuerte pueden tirar al arco |
| Jumping | Cuán alto salta |
| Stamina | Resistencia a cansancio, lesiones y recuperación |
| Strength | Cuán fuerte es |
| LongShots | Precisión de tiro fuera del área |

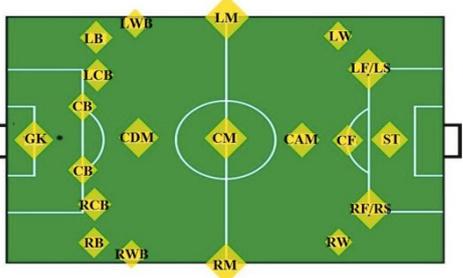
TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

| | |
|----------------|--|
| Aggression | Agresividad |
| Interceptions | Capacidad de "leer" el juego, anticipar pases |
| Positioning | Habilidad para posicionarse en el campo |
| Vision | Conciencia de sus compañeros de equipo |
| Penalties | Precisión en los penales |
| Composure | Compostura de un jugador cuando está bajo presión de un oponente |
| Marking | Marca |
| StandingTackle | Quite parado |
| SlidingTackle | Quite en barrida |
| GK Diving | Atajadas/Voladas |
| GK Handling | Capacidad de atrapar el balón sin rebotes |
| GK Kicking | Precisión en saque de meta |
| GK Positioning | Posicionamiento |
| GK Reflexes | Reflejos |

La mayor parte de los atributos mencionados presentan una escala de valores de 1 a 100. Debido a que todos los atributos deberán ser transformados a una misma escala, para algunos de ellos se aplicaron las siguientes transformaciones:

| Atributo | Escala Original | Escala Transformada |
|-------------|---|--------------------------------------|
| Weak Foot | Valores de 1 a 5 | Valor original multiplicado por 20 |
| Skill Moves | Valores de 1 a 5 | Valor original multiplicado por 20 |
| Work Rate | Low = Bajo Medium = Medio High = Alto | Low = 30 Medium = 60 High = 90 |

TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

| | | |
|----------|--|---|
| Position | <p>La <i>Figura 1</i> indica la descripción de posiciones dentro del campo de juego:</p> <div style="text-align: center;">  </div> <p style="text-align: center;">Figura 1. Posiciones dentro del campo de juego.</p> | <p>GK = 1</p> <p>CB/LCB/RCB = 10</p> <p>RB = 20 LB = 25</p> <p>RWB = 30 LWB = 35</p> <p>CDM/LDM/RDM = 40</p> <p>CM/LCM/RCM = 50</p> <p>RM = 60 LM = 65</p> <p>CAM/RAM/LAM = 70</p> <p>RW/LW/CF = 80</p> <p>RF/RS = 85</p> <p>LF/LS = 90</p> <p>ST = 100</p> |
|----------|--|---|

Por otro lado, los atributos seleccionados para presentar el resultado de los futbolistas seleccionados como similares son los siguientes:

| Atributos | Descripción |
|----------------|------------------------|
| Name | Nombre |
| Age | Edad |
| Nationality | Nacionalidad |
| Club | Club actual |
| Value | Valor del pase |
| Wage | Salario |
| Release Clause | Cláusula de Liberación |

9.2 Fase II

Objetivo: Aplicar componentes principales a los atributos existentes en la base de datos.

Para reducir la cantidad de variables dentro del cálculo matemático que determina el nivel de similitud entre los futbolistas, se aplicó el concepto de componentes principales. Previo a implementar componentes principales (PCA), se han ponderado aquellos atributos que se consideran más relevantes según la posición que ocupa un jugador dentro del campo de juego. Es decir, si el futbolista en cuestión es un defensor se ponderarán atributos futbolísticos distintos a los de un delantero. Por este motivo a la hora de implementar componentes principales (PCA) se optó por

TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

utilizar la matriz de covarianza, ya que otorga mayor peso a los atributos que presentan mayor variabilidad.

Como resultado, se obtuvieron 31 nuevos atributos de 40 originales que conservan el 97% de variabilidad de los datos originales y no poseen correlación entre sí.

9.3 Fase III

Objetivo: Formar vectores que representen matemáticamente las componentes principales obtenidas para cada futbolista.

Para conformar vectores que representen las características de juego de los futbolistas, se utilizaron los valores de las primeras 31 componentes principales (97% de variabilidad de los datos originales).

Por lo tanto el proceso de armado de vectores es de la siguiente manera: el futbolista "A" inicialmente tiene valorados 40 atributos en la base de datos original. Luego, dependiendo la posición en el campo de juego del jugador "A", se ponderan determinados atributos. Es decir, se multiplica el valor de ciertos atributos en función de si "A" es arquero, defensor, delantero, etc. Posteriormente, se aplica componentes principales y se seleccionan las primeras 31 componentes obteniendo el 97% de variabilidad de los datos originales. Estos 31 valores son los que finalmente conforman el vector que representa las características de juego del futbolista "A".

9.4 Fase IV

Objetivo: Desarrollar un cálculo matemático cuyo resultado sea el nivel de similitud entre dos futbolistas.

Dado un conjunto de valores referidos a características técnicas de dos jugadores, existen múltiples cálculos posibles para definir el nivel de similitud entre ellos. En el presente trabajo de investigación se optó por llevar a cabo una serie de pasos que finalmente permitirá conocer la relación de similitud entre 2 futbolistas:

Inicialmente, se parte de una base de datos con la valoración de 40 atributos futbolísticos para cada jugador. Según la posición que ocupa dentro del campo de juego el futbolista a analizar, se multiplican por un *factor* (para el presente trabajo *factor*=10) los atributos que estén más relacionados con aquella posición. Por ejemplo si el futbolista a analizar es delantero, se multiplican por el *factor* los valores de *finishing*, *volleys*, *dribbling*, *agility* y *shotpower*. De esta forma se logra mayor variabilidad en los atributos más relevantes.

Luego se aplica el concepto de componentes principales. El objetivo es reducir la cantidad de variables a considerar y eliminar la correlación entre variables. Dado que los atributos que han sido multiplicados por el *factor* presentan mayor variabilidad, adquieren mayor relevancia a la hora de implementar componentes principales utilizando la matriz de covarianza.

Más tarde se conforman los vectores que representan a cada futbolista utilizando los valores de las primeras n (para el presente trabajo $n=31$) componentes principales. Por último se calcula la distancia euclidiana entre el vector del futbolista a analizar y todos los vectores de los jugadores restantes. Finalmente los 3 futbolistas que presentan la menor distancia euclidiana respecto del futbolista que se requiere analizar serán considerados los 3 colegas más similares.

La *Figura 2*, indica cómo se calcula la distancia euclidiana entre los puntos $P = (p_1, p_2, \dots, p_n)$ y $Q = (q_1, q_2, \dots, q_n)$ en el espacio n -dimensional.

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

Figura 2. Fórmula utilizada para calcular la distancia euclidiana en espacio n -dimensional.

9.5 Fase V

Objetivo: Ejecutar el algoritmo desarrollado utilizando casos reales.

A continuación se presentan los resultados obtenidos tras ejecutar el algoritmo desarrollado para detectar los 3 futbolistas más similares a partir de un jugador dado.

Para cada uno de los futbolistas encontrados se muestra su nombre (*Name*), edad (*Age*), nacionalidad (*Nationality*), club actual (*Club*), valor del pase (*Value*), salario anual (*Wage*) y cláusula de rescisión de contrato (*Release Clause*).

- Lionel Messi

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|-------------|------------|--------------------|---------------------|--------------|-------------|-----------------------|
| L. Messi | 31 | Argentina | FC Barcelona | €110.5M | €565K | €226.5M |
| E. Hazard | 27 | Belgium | Chelsea | €93M | €340K | €172.1M |
| Neymar Jr | 26 | Brazil | Paris Saint-Germain | €118.5M | €290K | €228.1M |
| P. Dybala | 24 | Argentina | Juventus | €89M | €205K | €153.5M |

TRABAJO FINAL INTEGRADOR – Jugadores de Fútbol Similares

- Ángel Dí María

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|-------------|-----|-------------|---------------------|--------|-------|----------------|
| A. Di María | 30 | Argentina | Paris Saint-Germain | €29.5M | €125K | €54.6M |
| T. Lemar | 22 | France | Atlético Madrid | €35.5M | €64K | €79.9M |
| E. Lamela | 26 | Argentina | Tottenham Hotspur | €21M | €105K | €40.4M |
| Lucas Moura | 25 | Brazil | Tottenham Hotspur | €31M | €115K | €59.7M |

- Neymar Jr

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|------------|-----|-------------|---------------------|---------|-------|----------------|
| Neymar Jr | 26 | Brazil | Paris Saint-Germain | €118.5M | €290K | €228.1M |
| E. Hazard | 27 | Belgium | Chelsea | €93M | €340K | €172.1M |
| D. Mertens | 31 | Belgium | Napoli | €45M | €135K | €76.5M |
| L. Messi | 31 | Argentina | FC Barcelona | €110.5M | €565K | €226.5M |

- Luis Suárez

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|----------------|-----|-------------|---------------------|--------|-------|----------------|
| L. Suárez | 31 | Uruguay | FC Barcelona | €80M | €455K | €164M |
| H. Kane | 24 | England | Tottenham Hotspur | €83.5M | €205K | €160.7M |
| E. Cavani | 31 | Uruguay | Paris Saint-Germain | €60M | €200K | €111M |
| R. Lewandowski | 29 | Poland | FC Bayern München | €77M | €205K | €127.1M |

- De Gea

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|---------------|-----|-------------|-------------------|--------|-------|----------------|
| De Gea | 27 | Spain | Manchester United | €72M | €260K | €138.6M |
| K. Schmeichel | 31 | Denmark | Leicester City | €19M | €78K | €36.1M |
| Ederson | 24 | Brazil | Manchester City | €41.5M | €125K | €79.9M |
| M. Neuer | 32 | Germany | FC Bayern München | €38M | €130K | €62.7M |

- Thiago Silva

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|---------------|-----|-------------|---------------------|--------|-------|----------------|
| Thiago Silva | 33 | Brazil | Paris Saint-Germain | €24M | €165K | €44.4M |
| S. Umtiti | 24 | France | FC Barcelona | €57M | €205K | €121.1M |
| J. Vertonghen | 31 | Belgium | Tottenham Hotspur | €34M | €155K | €62.9M |
| N. Aké | 23 | Netherlands | Bournemouth | €14.5M | €55K | €28.6M |

- Cristiano Ronaldo

| Name | Age | Nationality | Club | Value | Wage | Release Clause |
|-------------------|-----|-------------|---------------------|--------|-------|----------------|
| Cristiano Ronaldo | 33 | Portugal | Juventus | €77M | €405K | €127.1M |
| S. Agüero | 30 | Argentina | Manchester City | €64.5M | €300K | €119.3M |
| R. Lewandowski | 29 | Poland | FC Bayern München | €77M | €205K | €127.1M |
| K. Mbappé | 19 | France | Paris Saint-Germain | €81M | €100K | €166.1M |

9.6 Fase VI

Objetivo: Validar resultados obtenidos

Verificando los resultados obtenidos en la sección anterior, se puede notar que en la totalidad de los casos se han detectado jugadores realmente similares entre sí. Incluso las posiciones que ocupan dentro del campo de juego son las esperadas. El ejemplo más notorio fue “De Gea”. Es arquero y todos sus jugadores similares también ocupan esta misma posición.

Se puede concluir que el algoritmo desarrollado detecta aquellas características más importantes de un futbolista y a partir de ellas logra encontrar exitosamente jugadores realmente similares a uno dado.

Link a la solución del trabajo realizado:

github.com/fcirigliano/jugadores_similares/blob/main/Jugadores_Similares.ipynb

10. DISCUSIÓN

Observando los resultados obtenidos se puede determinar que el algoritmo ha logrado resolver la problemática planteada inicialmente. Como aspectos positivos de la solución se destaca la diversidad de valor de mercado y salario en los jugadores resultantes como similares. Es decir, se amplía el espectro a la hora de contratar futbolistas que cumplan con determinadas cualidades. Ese es el diferencial de esta propuesta. Su validez es tanto para aquellos que pueden acceder al mercado de jugadores de fútbol más consagrados como para quienes necesitan contratar futbolistas que cumplan determinados requisitos pero que no pueden acceder al mercado de pases de las grandes estrellas del fútbol mundial.

Siempre es subjetivo determinar si un futbolista es realmente similar a otro, pero se puede observar que en los resultados obtenidos los protagonistas continúan una misma línea futbolística. Es claro que el algoritmo desarrollado ofrece una nueva herramienta para el área de contrataciones de un club, ya sea de los más poderosos o de los que menos recursos económicos tienen. Quedará a criterio de cada club en qué futbolistas similares posarán sus ojos. En la mayoría de los casos esta etapa de descubrimiento de jugadores debe ser continuada por una etapa que contemple un seguimiento minucioso de los profesionales que han despertado interés.

Para finalizar esta sección, se proponen nuevas líneas de investigación que podrían mejorar aún más la precisión de los resultados obtenidos. Una alternativa que se plantea es enriquecer aún más la base de datos para contemplar nuevas características relacionadas al aspecto físico de los jugadores. Por ejemplo, se podrían considerar kilómetros recorridos de cada jugador por partido y en qué sectores del campo de juego se distribuyeron. De esta forma se podrían determinar jugadores similares contemplando tanto los aspectos futbolísticos como los físicos.

11. CONCLUSIONES

Analizando los resultados obtenidos durante la investigación se puede afirmar que el algoritmo desarrollado logró comparar y valorar correctamente las características futbolísticas de los jugadores permitiendo detectar de esta forma profesionales similares entre sí.

Al tener en cuenta futbolistas de todas las ligas profesionales del mundo se considera a esta herramienta con un alto potencial y de suma utilidad para detectar futbolistas similares en la etapa de reclutamiento. Además, al realizar distintos análisis comparativos sin considerar el valor de mercado de los futbolistas, el algoritmo presentado podría ser utilizado tanto por clubes que poseen grandes presupuestos para reforzar su plantel como por aquellos clubes que no cuentan con dichos recursos. De esta forma se convierte en una gran oportunidad para clubes que pretenden futbolistas acordes a su presupuesto y con características futbolísticas similares a las grandes figuras del fútbol mundial.

12. REFERENCIAS BIBLIOGRÁFICA

- [1] S. Mclean, P. M. Salmon, A. D. Gorman, N. J. Stevens, and C. Solomon, “A social network analysis of the goal scoring passing networks of the 2016 European Football Championships,” *Hum. Mov. Sci.*, vol. 57, pp. 400–408, Feb. 2018, doi: 10.1016/j.humov.2017.10.001.
- [2] C. S. Loh, I. H. Li, and Y. Sheng, “Comparison of similarity measures to differentiate players’ actions and decision-making profiles in serious games analytics,” *Comput. Human Behav.*, vol. 64, pp. 562–574, Nov. 2016, doi: 10.1016/j.chb.2016.07.024.
- [3] J. Mazurek, “Which Football Player Bears Most Resemblance to Messi? A Statistical Analysis,” Feb. 2018, Accessed: Jul. 28, 2020. [Online]. Available: <http://arxiv.org/abs/1802.00967>.
- [4] R. Vroonen, T. Decroos, J. Van Haaren, and J. Davis, “Predicting the Potential of Professional Soccer Players.”
- [5] J. L. Peña and R. S. Navarro, “Who can replace Xavi? A passing motif analysis of football players,” Jun. 2015, Accessed: Jul. 28, 2020. [Online]. Available: <http://arxiv.org/abs/1506.07768>.
- [6] W. S. Erdmann and R. Dargiewicz, “FOOTBALL (SOCCER) PLAYERS KINEMATICS AT DIFFERENT DEVELOPMENT LEVELS. PART ONE: ACQUISITION AND PROCESSING OF DATA,” 2006. Accessed: Jul. 28, 2020. [Online]. Available: <https://ojs.ub.uni-konstanz.de/cpa/article/view/308>.
- [7] “FIFA player ratings explained: How are the card number & stats decided? | Goal.com.” <https://www.kaggle.com/karangadiya/fifa19> (accessed Aug. 09, 2020).