
ANÁLISIS MULTIVARIADO Y DE REGRESIÓN LOGISTICA PARA LA DETECCIÓN DE JUGADORES PROMESAS DEL FUTBOL

MULTIVARIATE AND LOGISTIC REGRESSION ANALYSIS FOR THE DETECTION OF PROMISING SOCCER PLAYERS

German Santiago Garzón Mora.^a
german.garzon@usantotomas.edu.co

Mario José Pacheco López.^b
mariopacheco@usantotomas.edu.co

1. Resumen

El Fútbol al ser el deporte más visto y practicado en el mundo, se ha convertido en uno de los negocios más rentables en los últimos años, debido a la gran cantidad de ingresos recibidos por parte de patrocinio de marcas, derechos televisivos, venta de jugadores entre muchos más contratos exclusivos para el “Deporte Rey”, todo esto ha traído como consecuencia que los jugadores de futbol con rango de edad entre los 22 años y 30 años se han vuelto muy costosos para cualquier equipo debido por este marketing que se ha realizado, por lo cual ha generado el interés en los clubes de futbol por identificar jugadores jóvenes que no son reconocidos por su corta carrera profesional pero que tengan habilidades físicas que se puedan potencializar en un futuro para formar un jugador con excelentes condiciones, todo esto con la finalidad de comprar al jugador a un precio muy económico y que en el transcurso de los años este jugador, a medida que aumenta sus habilidades, pueda generar mayores ingresos al club por medio de su venta a otro equipo o conceptos de marketing. Con el fin de poder identificar jugadores jóvenes con habilidades para el futbol, se propone implementar un modelo logístico para poder detectarlos, empleando información de la página SOFIFA (<https://sofifa.com/>) que contiene la información de más de 10.000 futbolistas profesionales, los datos son completamente abiertos y están disponibles por año, entonces se evidencia la evolución de los jugadores de las cualidades físicas y mentales de los jugadores con un total de 39 variables entre las que se encuentran la capacidad de Pases, la velocidad, la capacidad de disparo, la fuerza, el control de la pelota, la visión entre otras.

2. Introducción

Sánchez (2011) el Fútbol es un deporte de equipo jugado entre dos conjuntos de once jugadores cada uno y algunos árbitros que se ocupan de que las normas se cumplan correctamente, se juega en un terreno rectangular de césped natural o artificial, con una portería o arco a cada lado del campo, se juega con una pelota que se debe desplazar a través del campo con cualquier parte del cuerpo que no sean los brazos o las manos, y mayoritariamente con los pies, el objetivo del juego es introducirla dentro de la portería o arco contrario que es la acción de anotar un gol, al final el equipo que logre más goles en 90 minutos de juego más tiempo de reposición es el ganador del partido.

^aEstudiante

^bDirector

Meneses y Avalos (2013) describe que en 1884 se funda la Asociación Inglesa de Fútbol, que recoge todas las normas de este deporte, pero muchos siglos antes ya se habían creado en las islas británicas durante la edad media y desde el siglo III a.C juegos de pelota similares.

Con la realización de la primera reunión de la International Football Association Board en 1886 y la fundación de la FIFA en 1904, este deporte se ha expandido hasta llegar a todos los rincones del mundo. A partir de 1930 se comenzaría a disputar la Copa Mundial de Fútbol, que se convertiría en el evento deportivo con mayor audiencia del mundo, además en ese año también se dejó el juego brutal y callejero y paso a ser un juego de posesión de balón, ya que por primera vez se jugaba muy parecido al fútbol moderno con la diferencia de que la cantidad de defensas o atacantes pueden variar según el sistema de juego del equipo, Meneses y Avalos (2013) .

Para los mundiales de 1934, 1939 y 1950 los equipos formaron sistemas de juego basados en la preparación atlética, en 1970 el fútbol ya tenía formaciones establecidas que repartían a todos sus jugadores por el campo de juego con labores distintas, pero en 1974 hubo un sistema de juego que revolucionó, que fue el Fútbol Total creado por los fundamentos del inglés Jack Reynolds y fue perfeccionado por el Holandés Rinus Michels durante su estancia en el equipo Ajax de Ámsterdam y su auge fue en el mundial de 1974 con la selección de Holanda contando con el histórico jugador Johan Cruyff como jugador principal del novedoso sistema de juego, en el Fútbol Total consiste en que el jugador que se mueve fuera de su posición es sustituido por un compañero de equipo, lo que permite que la formación del conjunto siga con su estructura táctica, ningún futbolista tiene una posición fija y también este sistema implementó que el equipo hiciera una presión fuerte en la cancha del rival con el objetivo de recuperar el balón lo más rápido posible y tener al rival alejado del arco, Meneses y Avalos (2013) .

En las décadas de los 80 y 90 el fútbol tuvo más interés en los países tercermundistas, específicamente en África, ya que los jugadores de esta región tenían una capacidad física muy buena por delante de los jugadores europeos, también los países suramericanos empezaron a destacar por encima de los países europeos, el ejemplo de ello fue el campeonato mundial que ganó Argentina en 1986, también aumentaron la compra de los jugadores suramericanos por parte de los equipos europeos.

En los últimos años el fútbol se ha centrado en que los jugadores deben tener una elevada preparación física y en conocimientos de tácticas, volviendo al jugador no solo a futbolista sino a atleta, que tenga la capacidad de correr todo un partido sin cansarse pero cumpliendo las funciones que le asignó el entrenador, en la parte de las reglas se han realizado muchas investigaciones y debates que han generado modificación de las mismas con el objetivo de que el juego sea más justo y que se cuide la integridad física del jugador, adicional a ello desde el 2018 se implementó un nuevo sistema que apoya a los árbitros que es el VAR (Video Assistant Referee) que son una serie de cámaras que sigue todo el partido y que en caso de que el árbitro tenga una duda con una decisión que haya tomado, él puede solicitar revisar las cámaras y verificar su decisión.

Con respecto a estudios de rendimiento de futbolistas, Hirose, 2011 realiza estudios para la predicción de talento en futbolistas juveniles; estudio prospectivo durante 4-6 años: en el estudio participaron 62 jugadores de fútbol adolescentes. Se dividieron en jugadores profesionales, universitarios y regionales en función de su nivel de desempeño a los 18 años. Se midieron la altura, el peso, la edad esquelética, la velocidad de paso, la resistencia y el tiempo de reacción de elección (mano: HRT, pie: FRT y complejo mano-pie: CRT). CRT fue significativamente más rápido en jugadores profesionales y jugadores universitarios en comparación con los jugadores regionales. Los valores medios de peso corporal, estatura y TRC tendían a ser más altos para los jugadores profesionales que para los jugadores universitarios, aunque estadísticamente insignificantes. Nuestros hallazgos sugieren que los jugadores de fútbol adolescentes con TRC más rápido deberían identificarse como posibles jugadores profesionales, todos los datos se expresaron como media $\pm$ SD. Las diferencias en todas las mediciones entre los grupos en el nivel de desempeño

se evaluaron mediante ANOVA de un factor y la prueba de diferencia mínima significativa de Scheffé.

(Vroonen, Decroos y Van Haaren 2016) en su artículo Predicting the Potential of Professional Soccer Players, SciSports presentan un sistema de proyección para futbolistas llamado APROPOS que está inspirado en los sistemas CARMELO y PECOTA. APROPOS predice el potencial de un jugador de fútbol al buscar en una base de datos histórica para identificar jugadores similares de la misma edad. Luego basa su predicción para la progresión del jugador objetivo en cómo evolucionaron realmente los jugadores anteriores similares. El algoritmo que se diseñó se llama APROPOS (Algoritmo de predicción del potencial de los jugadores de fútbol). Al igual que los sistemas de proyección PECOTA y CARMELO, utiliza un enfoque de vecinos más cercanos para predecir cómo evolucionará la habilidad de un jugador de fútbol con el tiempo.

(Razali y Mustapha 2017) desarrollaron un trabajo con el objetivo de introducir un nuevo marco desde la perspectiva de la informática para identificar talentos en el deporte del fútbol basado en las cualidades individuales de los jugadores; físico, mental y técnico. La combinación de cualidades evaluadas por los entrenadores se utiliza para predecir la posición de los jugadores en un partido que se adapta mejor al jugador en una formación de equipo en particular. La evaluación del marco propuesto es doble; cuantitativamente mediante experimentos de clasificación para predecir la posición del jugador, y cualitativamente a través de un sitio de identificación de talentos desarrollado para lograr el mismo objetivo. Los resultados de los experimentos de clasificación utilizando redes bayesianas, árboles de decisión y K-vecino más cercano han mostrado un promedio de 98 % de precisión, lo que promoverá la coherencia en la toma de decisiones mediante la eliminación del sesgo personal en la selección del equipo. Este artículo propuso una metodología empírica que utiliza un enfoque de minería de datos para medir cuantitativamente las fortalezas de cada jugador en diferentes posiciones en función de sus habilidades físicas, mentales y técnicas en el fútbol.

(Jauhainen 2019), realiza una investigación en la que se aborda la identificación del talento como un problema de detección de anomalías. Entrenamos una máquina de vectores de apoyo no lineal de una clase (SVM de una clase) en un conjunto de datos ($N = 951$) recopilado de jugadores de fútbol junior de 14 años para detectar posibles futuros jugadores de élite. El área media bajo la curva característica de funcionamiento del receptor (AUC-ROC) sobre las combinaciones de hiperparámetros probadas fue 0,763 (estándar 0,007). El modelo más preciso se obtuvo cuando se utilizaron pruebas físicas, midiendo, por ejemplo, habilidades técnicas, velocidad y agilidad. De acuerdo con nuestros resultados, el enfoque propuesto podría ser útil para apoyar a los tomadores de decisiones en el proceso de identificación del talento, se entrena el support vector machine de una clase (SVM de una clase) (Chandola, Banerjee y Kumar, 2009) para modelar la región normal de los datos basados solo en las observaciones de la clase mayoritaria, es decir, la no - Jugadores de la academia. El modelo entrenado se puede utilizar para predecir si las nuevas observaciones pertenecen a esta región normal o no.

Como se mencionaba anteriormente, el marketing en el fútbol ha aumentado las ganancias en este deporte, Gerhard (1974) desde su óptica de sociólogo marxista, hizo varios señalamientos acertados y vigentes sobre los problemas económicos que amenazaban al fútbol, y el tiempo le ha dado la razón. Señaló que el endeudamiento de los clubes de fútbol es consustancial a la lógica del sistema capitalista. Lo que a su vez explica que la veta de negocio televisión/fútbol haya sido exprimida ininterrumpidamente y siga rindiendo beneficios. Vinnai (1974) ya dijo que los clubes de jugadores profesionales son empresas económicas profesionales del sector de la prestación de servicios, que venden las exhibiciones de futbolistas, en cuanto mercancía, a un público que las consume. Hoy esto es más que evidente y los clubes de fútbol profesional solo son viables como actividad empresarial, en la medida que sean capaces de reducir la actividad deportiva a la condición de una mercancía exitosa deportiva y comercialmente.

Para el presente trabajo se implementó un modelo de regresión logístico para detectar jugadores jóvenes

con buenas habilidades que sean a futuro buenos futbolistas.

La regresión logística es uno de los instrumentos estadísticos más versátiles de que se dispone para el análisis de datos, su origen se remonta en 1961 (Confield, Gordon y Smith 1961); su uso se expande desde principios de los ochenta por las facilidades informáticas que se contaban desde entonces, en los últimos años se ha utilizado mucho esta técnica, tanto en la literatura orientada a tratar temas metodológicos como en los artículos científicos biomédicos.

3. Problema

En la detección de jugadores jóvenes potencialmente buenos a futuro por medio de análisis estadístico se tienen diferentes antecedentes que han arrojado modelos que no tienen en cuenta muchas variables físicas del jugador y también no tienen en cuenta la evolución año tras año de un futbolista profesional para poderlo comparar con un jugador joven que podría ser una nueva promesa de este deporte.

4. Pregunta Problema

¿Es posible detectar promesas en el futbol por medio de predicciones de un modelo de regresión logística y que arroje resultados buenos?

5. Objetivo General

- Ajustar un modelo de regresión logística para la detección de jugadores promesas del futbol.

6. Objetivo Específicos

- Identificar las variables que hacen que un futbolista se destaque por encima de otros.
- Categorizar los tipos de futbolistas según las habilidades que tiene para poderlos clasificar en los distintos sistemas de juego que tiene el futbol.

7. Justificación

El futbol es el deporte más famoso en el mundo, por su sistema de juego entretenido ha generado una gran cantidad de seguidores que lo practican y lo visualizan, es tanto la fanaticada por este deporte que los eventos televisivos más vistos en el mundo en el transcurso de la historia han sido partidos de futbol, con el paso del tiempo el tiempo se ha combinado el futbol con el auge del análisis de datos que ha permitido formar bases de datos muy sólidas con información de resultados de partidos, cualidades de cualquier futbolista profesional, estado físico de un futbolista, cantidad de tiros en un partido, cantidad de faltas en una temporada entre muchos datos que al momento de traducir esta información por medio

de análisis estadístico podría ser de gran ayuda para la mejora de este deporte, esta combinación es tan innovadora tanto así que los protagonistas del juego, es decir los equipos de futbol están entrando en esta ola del análisis de datos para dejar la evaluación visual tradicional y comenzar con un análisis de datos y les permita mejorar sus procesos a un mejor costo y a conformar buenos equipos de juego para el futuro; el futbol además de ser el deporte más visto en el mundo, es uno de los negocios que más dinero mueven, a tal grado que los mejores futbolistas están entre la lista de los más millonarios en el mundo, pero para poder llegar a ser un buen jugador se necesita de muchas cualidades físicas y mentales que todos los días deben mejorar, y esto es lo que hace que cuando un jugador está en un punto muy bueno de su carrera deportiva valga mucho dinero para un equipo de futbol y esto en uno de los pilares para los equipos, poder detectar jugadores que tengan una buena proyección en el futbol para el futuro y lo puedan comprar a un precio razonable para poder mejorar sus habilidades y en un futuro poder venderlo por un valor mucho más alto por el que se compró inicialmente o también poderlo mantener para conformar un buen equipo.

Para solucionar este problema los equipos contratan ojeadores de futbol que van por todo el mundo detectando talentos que les pueda traer una buena contribución al equipo de interés, pero esta detección de talentos es muy visual del ojeador y no cuenta con un informe que le de las cualidades del jugador en estudio y esto puede traer el problema de que se están seleccionando jóvenes que no van a poder llegar a la cima del futbol de elite y que en otros casos los verdaderos talentosos jugadores se están omitiendo, por eso esté presente trabajo de investigación se busca proponer un modelo logístico múltiple para detectar futuras promesas del futbol, los datos a utilizar son de la página SOFIFA que contiene la información de más de 10.000 futbolistas profesionales, los datos fueron tomados del video juego FIFA que son completamente abiertos y tiene información recolectada por año, evidenciando la evolución de los jugadores año tras año, las variables que nos suministra la página SOFIFA son valoraciones de las cualidades físicas y mentales de los jugadores que en total son 36 variables entre las más importantes son Pases, velocidad, disparo, fuerza, control de pelota, visión entre otras que reflejan por puntaje sus capacidades, el modelo propuesto es muy fuerte en la clasificación y en las predicciones que se destacan por encima de otros modelos estadísticos ya que en otros estudios de otras temas distinto al deporte han dado predicciones muy acertadas con errores muy pequeños.

En la parte social el presente trabajo va ayudar a contribuir al deporte en encontrar más deportistas de una manera más optima de como se viene haciendo para el caso de Colombia que es muy visual y no tienen respaldo en los datos, desafortunadamente en Colombia el deporte no se considera una carrera profesional, a lo cual para los deportistas en formación les cuesta mucho llegar a ser atletas profesionales y las únicas opciones que tienen es desertar del deporte y dedicarse a otra carrera o conseguir un patrocinador que les cobra dinero para que el por medio de sus influencias lo conviertan en un profesional en el deporte, lo cual es tema de corrupción que ocurre actualmente en el deporte; por otro lado una de las formas de combatir la delincuencia y las drogas es por medio del deporte y si se identifican más deportistas se puede reducir esa problemática social que tiene afectado a nuestro país.

8. Hipótesis

Nula: Es posible identificar jugadores promesas del Fútbol por medio de un modelo de regresión logística.

Alternativa: No es posible identificar jugadores promesas del Fútbol Por medio de un modelo de regresión logística.

9. Marco teórico

9.1. El Fútbol

Es un deporte en donde compiten dos equipos formados por once jugadores cada uno, regidos por un árbitro que es el encargado de que se cumplan las reglas del juego, el encuentro se realiza en un terreno de forma rectangular de césped artificial o natural, en cada uno de los lados del campo se encuentra un arco que está protegida por el arquero que es uno de los jugadores del equipo.

Este deporte consiste en que un balón debe ser transportado por los equipos a través de todo el terreno de juego, para ello pueden utilizar cualquier parte del cuerpo (normalmente por los pies) a excepción de los brazos y las manos, el único jugador que puede tomar el balón con las manos es el arquero, siempre y cuando se encuentre dentro de su área o portería y no incumpla con las normas establecidas.

El objetivo del juego consiste en introducir el balón dentro de la portería del equipo contrario, esta acción se conoce como marcar un gol, la duración del partido es de 90 minutos repartido en dos partes de 45 minutos y cada tiempo puede tener tiempo de reposición que es el tiempo compensado cuando el balón no rodo durante el tiempo normal reglamentario (como por ejemplo faltas, lesiones, verificación de decisiones arbitrales, entre otros), al final del encuentro el equipo que haya marcado más goles es el ganador del partido.

El futbol 11 se rige por 17 normas que se usan de manera universal, pero algunas pueden ser modificadas en el futbol femenino, infantil o juvenil para mejorar su desarrollo y por falta de algunas capacidades que no tienen aún, como por ejemplo en el futbol infantil la duración del encuentro mucho menor ya que los niños no tienen aún la capacidad física para aguantar un partido de 90 minutos, también las aplicaciones de estas reglas pueden ser diferentes según la región futbolística en la cual se está realizando el encuentro ya que pueden existir variaciones por múltiples aspectos.

El terreno donde se realizan los partidos, como se mencionaba anteriormente debe ser rectangular y normalmente debe tener medidas entre 90 y 1,20 metros de largo y de 45 a 90 metros de ancho, en partidos internacionales estas medidas pueden cambiar; el campo de juego está delimitado por líneas laterales o de banda y las que están a lo ancho son las líneas de meta, adicional a ello dentro del campo también se encuentran unas zonas delimitadas por líneas que son la portería que es donde el arquero puede tomar el balón con las manos y el círculo central que es donde se hacen los saques iniciales o después de marcar un gol.

Cada equipo para temas de orden en su sistema de juego cuenta con posiciones que cumplen funciones específicas durante el encuentro, el guardameta se encarga de cuidar la portería con el fin de evitar que el equipo rival marque un gol, puede utilizar las manos para tapar el balón y es el único que lo puede hacer, en cada equipo solo puede haber un arquero.

Los defensas es el jugador que se encuentra por delante del arquero y su función es frenar todos los ataques del equipo contrario, existen tipos de defensas como el defensa central que es el que juega en la parte de abajo en el centro que tiene normalmente la cualidad de ser un jugador muy alto, fuerte y cabecea muy bien pero tiene la desventaja en ser un jugador lento y no es tan hábil con el balón; también está el defensa lateral que son los encargados de defender por las bandas del campo de juego, tienen la función de sacar al equipo para poder llegar al arco del rival, estos jugadores se caracterizan por ser muy rápidos y con una buena condición física.

El mediocampista juega en el mediocampo del terreno de juego, estos tienen muchas funciones como recuperar balones, creación de jugadas de gol y sacar provecho del juego defensivo, también existen tipos de mediocampistas que cumplen funciones específicas en la mitad del campo, como el volante de recuperación que se encarga de solo recuperar el balón en la mitad, estos jugadores tienen cualidades parecidas a los defensas centrales pero tienen mejor cualidad física, también están los volantes de creación que son los encargados de crear jugadas que terminen en gol, se caracterizan por ser jugadores muy hábiles con la pelota, velocidad mental para tomar decisiones y una visión del campo muy amplia, por último están los volantes mixtos que es la combinación de los dos anteriores mencionados, las habilidades de este jugador además de tener las habilidades juntas del volante de recuperación y creación, se destacan por tener una muy buena condición física ya que durante todo el partido debe recorrer toda la cancha a diferencia de otras posiciones.

El delantero es el encargado de anotar goles, está ubicado cerca de la portería del equipo rival y siempre está en posición de ataque, es el jugador más valorado del equipo porque es de los partícipes del objetivo del juego que es anotar el juego, son los más costosos para los equipos y de cierta manera un poco injusta son los más reconocidos en los partidos, se caracterizan por tener velocidad, fuerza y buen disparo.

Con respecto a las anteriores posiciones explicadas anteriormente se aclara que en un partido los jugadores pueden cumplir funciones de otras posiciones si lo requiere, por ejemplo, un defensor central también puede marcar un gol en un tiro de esquina o un delantero también puede bajar a ayudar a defender para que el otro equipo no convierta gol.

9.2. Modelo de regresión Logístico

El algoritmo de regresión logística es uno de los más utilizados hoy en día en aprendizaje automático, su principal aplicación los problemas de clasificación binaria. Es un algoritmo simple de fácil interpretación de resultados que, aunque a pesar de su simplicidad funciona muy bien en las aplicaciones y se utiliza como referencia de rendimiento.

En la regresión logística se mide la relación entre la variable dependiente, la afirmación que se desea predecir, con una o más variables independientes, el conjunto de características disponibles para el modelo. Para esto se utiliza una función logística que arroja la probabilidad de la variable dependiente, el objetivo del modelo es buscar una clasificación, por lo que la probabilidad se debe traducir en valores binarios, para evaluar si la probabilidad es alta se define un umbral según el contexto del problema, generalmente el umbral es de 0,5 donde si está por encima de este umbral la afirmación es cierta y lo contrario es falsa la afirmación.

La función logística que relaciona la variable dependiente con las dependientes es la función sigmoidea, que es una curva en forma de S que puede tomar valores entre 0 y 1, la ecuación es:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i$$

$$\text{logit}(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i$$

El valor de la probabilidad de Y se puede obtener con la inversa del logaritmo natural:

$$p(Y) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i}}$$

En donde x es un número real; en la ecuación x cuando tiene a menos infinito el cociente tiende a cero y cuando tiene a infinito el cociente tiende a uno.

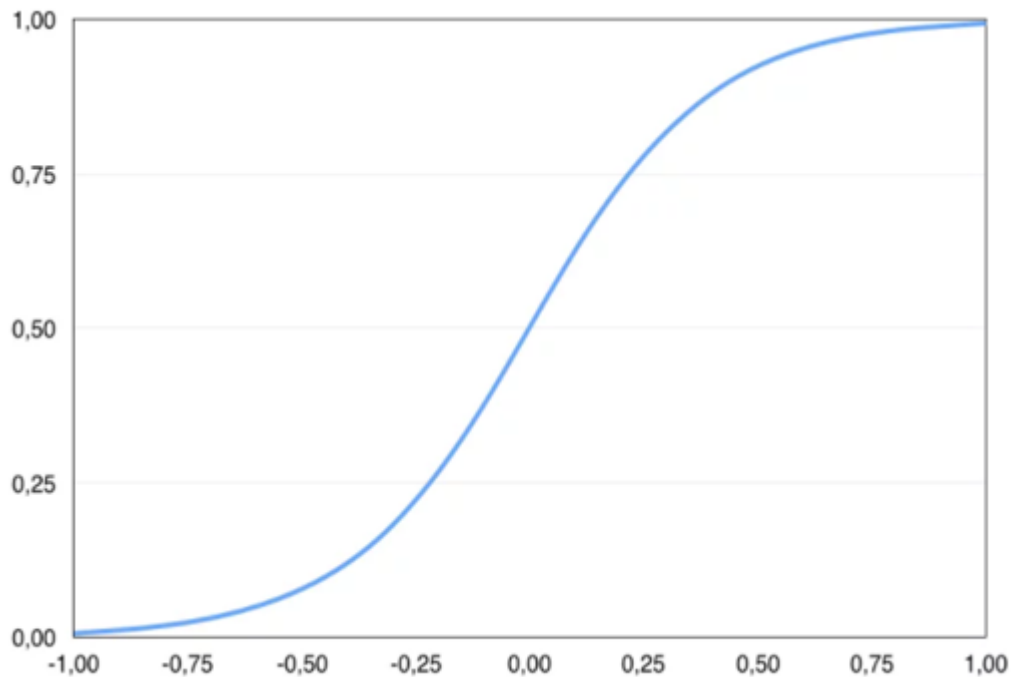


Figura 1: Gráfico de regresión logística, Fuente: <https://www.analyticslane.com>

La regresión logística analiza los datos distribuidos de la forma:

$$Y_i \sim \beta(p_i, n_i), \text{ para } i = 1, \dots, m$$

- n_i son los números de ensayos Bernoulli conocidos.
- p_i son las ptobabilidades de éxito que son desconocidas.

El modelo es obtenido a base de lo que cada ensayo y el conjunto de variables explicativas/independientes puedan informar acerca de la probabilidad final.

Las variables explicativas pueden pensarse como un vector X_i k -dimensional y el modelo toma la siguiente forma:

$$p_i = E \left(\frac{Y_i}{n_i} \middle| X_i \right)$$

Los logits de las probabilidades binomiales desconocidas son modeladas como una función lineal de los X_i .

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

- Un elemento particular de X_i puede ser ajustado a 1 para todo i obteniéndose una constante independiente en el modelo.
- Los parámetros β_j son desconocidos y se estiman normalmente por medio de máxima verosimilitud.

El modelo tiene una formulación equivalente dada por:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}$$

El proceso para entrenar una función logística se puede realizar maximizando la probabilidad de que los puntos del conjunto de datos clasifiquen de manera correcta.

10. Metodología General

Inicialmente para las bases se utilizarían técnicas de minería de datos para extraer de la página SOFIFA información de los futbolistas de las mejores ligas del mundo, después ya contando con una base de datos se empieza a analizar la información obtenida.

Variables:

En la página SOFIFA cada jugador cuenta con una puntuación de 1 a 100 en cada cualidad, las variables a utilizar se encuentran por categorías que buscan medir la misma habilidad pero desde aspectos distintos por ejemplo las variables de movimiento miden en general la velocidad del jugador en general pero miden la velocidad en carrera, la velocidad de las reacciones que tiene; o las variables de habilidad miden la capacidad que tiene el jugador con el balón, pero para ello toman en cuenta el regate que tiene el jugador o la precisión de pase :

Variable	Significado
Sofifa id	Numero de identificación único del jugador asignado por sofifa.
Short name	Nombre del Jugador.
Age	Edad del jugador.
Height cm	Estatura del jugador.
Weight kg	Peso del jugador.
Nationality	Nacionalidad del Jugador.
Club name	Equipo del jugador.
League name	Liga donde juega el futbolista.
League rank	Ranking de la liga donde juega (1 a 5) donde 1 es el mas alto.
Overall	Calificación general del jugador.
Potential	Calificación esperada del jugador.
Value eur	Valor del jugador en Euros.
Wage eur	Sueldo del jugador en Euros.
Release clause eur	Clausula del jugador en euros..
Team position	Posicion del jugador.
Attacking crossing	Capacidad de cruzar el balón de un lado de la cancha a otro.
Attacking finishing	Capacidad de definición de balón.
Attacking heading accuracy	Puntería de disparo.
Attacking short passing	Pases en distancias reducidas.
Attacking volleys	Capacidad para realizar voleas.
Skill dribbling	Capacidad para evadir jugadores rivales.
Skill curve	Al disparar el balón que tanta toma curva.
Skill fk accuracy	Precisión en el pase.
Skill long passing	Capacidad de realizar pases a distancias alejadas.
Skill ball control	Capacidad que evalúa los desplazamientos con el balón.
Movement acceleration	Es el incremento de velocidad de un jugador.
Movement sprint speed	Mide la velocidad del jugador.
Movement agility	Mide que tan rápido es el jugador mientras se gira.
Movement reactions	Mide la velocidad de reacción.
Movement balance	Mide la capacidad de equilibrio del jugador.
Power shot power	Mide la fuerza del disparo del jugador.
Power jumping	Mide la capacidad de salto del jugador.
Power stamina	Mide la resistencia del jugador al chocar.
Power strength	Mide la fuerza del jugador.
Power long shots	Mide la capacidad de disparo desde distancias alejadas.
Mentality aggression	Mide la agresividad de un jugador.
Mentality interceptions	Mide la capacidad de cortar jugadas.
Mentality positioning	Mide el posicionamiento del jugador..
Mentality vision	Mide la capacidad del jugador en crear jugadas.
Mentality penalties	Mide la capacidad que tiene el jugador para cobrar penaltis.
Mentality composure	Mide el estado del jugador.
Defending standing tackle	Mide la efectividad de quitar el balón sin barrerse.
Defending sliding tackle	Mide la efectividad de quitar el balón barriéndose.
Goalkeeping diving	Mide la fuerza del jugador en las manos como arquero.
Goalkeeping handling	Mide la capacidad de agarre de balón como arquero.
Goalkeeping kicking	Mide la forma de patear el balón, pero como arquero.
Goalkeeping positioning	Mide el posicionamiento del jugador como arquero.
Goalkeeping reflexes	Mide los reflejos del jugador como arquero.
Goalkeeping handling	Mide la capacidad de agarre de balón como arquero.
Goalkeeping kicking	Mide la forma de patear el balón, pero como arquero.
Goalkeeping positioning	Mide el posicionamiento del jugador como arquero.
Goalkeeping reflexes	Mide los reflejos del jugador como arquero.

Tabla 1: Variables, Fuente: Elaboración propia.



Lionel Andrés Messi Cuccittini
 **RW ST CF** 33y.o. (Jun 24, 1987) 5'7" 159lbs

93
Overall Rating

93
Potential

€103.5M
Value

€560K
Wage

PROFILE
 Preferred Foot **Left**
 4 ★ Weak Foot
 4 ★ Skill Moves
 5 ★ International Reputatio...
 Work Rate **Medium/ Low**
 Body Type **Unique**
 Real Face **Yes**
 Release Clause **€212.2M**
 ID **158023**

PLAYER SPECIALITIES
 #Dribbler
 #Distance Shooter
 #FK Specialist
 #Acrobat
 #Clinical Finisher
 #Complete Forward

FC BARCELONA

84 ★★★★★
 Position **RF**
 Jersey Number **10**
 Joined **Jul 1, 2004**
 Contract Valid Until **2021**

ARGENTINA

83 ★★★★★
 Position **RW**
 Jersey Number **10**

LIKE (1304) DISLIKE (430) FOLLOW (1277) HISTORY VERSION (574) ▾

LAYOUT 1 2 3

ATTACKING
85 Crossing
95 Finishing
70 Heading Accuracy
91 Short Passing
88 Volleys

SKILL
96 Dribbling
93 Curve
94 FK Accuracy
91 Long Passing
96 Ball Control

MOVEMENT
91 Acceleration
80 Sprint Speed
91 Agility
94 Reactions
95 Balance

POWER
86 Shot Power
68 Jumping
72 Stamina
69 Strength
94 Long Shots

MENTALITY
44 Aggression
40 Interceptions
93 Positioning
95 Vision
75 Penalties
96 Composure

DEFENDING
32 Defensive Awareness
35 Standing Tackle
24 Sliding Tackle

GOALKEEPING
6 GK Diving
11 GK Handling
15 GK Kicking
14 GK Positioning
8 GK Reflexes

TRAITS
 Finesse Shot
 Long Shot Taker (AI)
 Speed Dribbler (AI)
 Playmaker (AI)
 Outside Foot Shot
 One Club Player

Figura 2: Visual página SOFIFA, Fuente: <https://sofifa.com>

Después de tener la base de datos, se realizó un análisis por componentes principales que me permita identificar variables similares y las más significativas y que me permita crear grupos o ejes con ciertas habilidades, también se implementó métodos multivariados de clasificación para lograr agrupar a todos los individuos de la base de datos y se tenga con más claridad las habilidades que tenga un grupo de jugadores.

En la siguiente fase se implementó el modelo que cumpla todos los supuestos y se va a ajustar el mejor modelo con las variables significativas para después comparar los resultados y poder definir cuál fue el mejor, por último, se sacan los resultados con sus respectivas conclusiones.

Para el presente trabajo se utilizarán tres softwares para conseguir los objetivos, el primero será Microsoft Excel para almacenar los datos descargados por la página SOFIFA, Python para realizar el procedimiento de Web Scraping y R studio para los análisis estadísticos y los modelos, en caso de que no se pueda realizar los modelos por R studio, se utilizará Python.

11. Resultados

Los datos fueron obtenidos de blogs que comparte la página SOFIFA, las bases obtenidas son totalmente abiertas, ya que el juego FIFA cada año revela las cualidades de los jugadores y están a disposición de cualquier jugador del videojuego.

Se obtienen en total 7 bases de datos desde el año 2015 hasta el año 2021 con la información de todos los jugadores de fútbol que están dentro del videojuego, cada base contiene aproximadamente más de 15.000 registros de jugadores de la mayoría de las ligas del mundo, por temas de derechos de ligas y equipos de fútbol existen años en que no salen en el videojuego, lo cual hace que se pueda no tener información de un mismo jugador durante los siete años, por ejemplo la liga Colombiana está desde el año 2015 hasta el año 2020, para el siguiente año por lo mencionado anteriormente no salieron todos los equipos de Colombia.

Para poderle dar un enfoque al trabajo se realiza una reducción en las bases de datos, tomando solo los jugadores latinoamericanos, se aclara que los jugadores de los países seleccionados no tienen la misma proporción en las bases de datos debido a los temas de derechos que se mencionaban anteriormente y por la importancia del fútbol en cada país, como por ejemplo la base de datos contiene más jugadores colombianos que cubanos ya que el videojuego FIFA no tiene la licencia de la liga cubana de fútbol, otro caso puede ser que las bases de datos tienen más jugadores argentinos que jugadores venezolanos ya que el fútbol en Argentina es el deporte más practicado, mientras que en Venezuela se tienen otros deportes más famosos; también por último de la base de datos se realizaron inclusiones o exclusiones de jugadores que tienen doble nacionalidad, para las bases de datos solo se tomó en cuenta el lugar de nacimiento del jugador, es decir si un jugador nació en Colombia pero adquirió la nacionalidad española, para la base de datos se toma el jugador para los siete años con nacionalidad colombiana.

Para poder proponer el modelo primero se vuelve a reducir la base de datos, tomando de los jugadores latinoamericanos únicamente los jugadores que salen durante los siete años, después se va a realizar un análisis de componentes principales en cada año para poder reducir variables de estudio que no aporten mucha información para poder realizar una predicción.

11.1. Análisis de componentes principales

El conjunto de datos de cada año contiene 592 individuos y 50 variables cuantitativas que miden capacidades físicas y mentales de jugadores de fútbol latinoamericanos.

11.1.1. Distribución de la varianza

Para el año 2021 las tres primeras dimensiones explican el 75,39 % de la varianza total del conjunto de datos; esto quiere decir que el 75,39 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 50,07 %, el segundo plano tiene aporta 15,01 % de la variabilidad total y el tercer plano aporta un 10.30 %.

Para el año 2020 las tres primeras dimensiones explican el 75,56 % de la varianza total del conjunto de datos; esto quiere decir que el 75,56 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 50,07 %, el segundo plano tiene aporta 14,83 % de la variabilidad total y el tercer plano aporta un 10.65 %.

Para el año 2019 las tres primeras dimensiones explican el 76,21 % de la varianza total del conjunto de datos; esto quiere decir que el 76,21 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 51,02 %, el segundo plano tiene aporta 14,63 % de la variabilidad total y el tercer plano aporta un 10.55 %.

Para el año 2018 Las tres primeras dimensiones explican el 76,02 % de la varianza total del conjunto de datos; esto quiere decir que el 76,02 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 51,36 %, el segundo plano tiene aporta 14,40 % de la variabilidad total y el tercer plano aporta un 10.25 %.

Para el año 2017 Las tres primeras dimensiones explican el 75,99 % de la varianza total del conjunto de datos; esto quiere decir que el 75.99 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 51,52 %, el segundo plano tiene aporta 14,40 % de la variabilidad total y el tercer plano aporta un 10.06 %.

Para el año 2016 Las tres primeras dimensiones explican el 76,81 % de la varianza total del conjunto de datos; esto quiere decir que el 76,81 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 51,78 %, el segundo plano tiene aporta 14,33 % de la variabilidad total y el tercer plano aporta un 10.70 %.

Para el año 2015 Las tres primeras dimensiones explican el 74,82 % de la varianza total del conjunto de datos; esto quiere decir que el 74,82 % de la variabilidad total de las variables se explica por estos planos. Este porcentaje es particularmente alto y, por lo tanto, el primer plano equivale al 48,59 %, el segundo plano tiene aporta 15,17 % de la variabilidad total y el tercer plano aporta un 11.04 %.

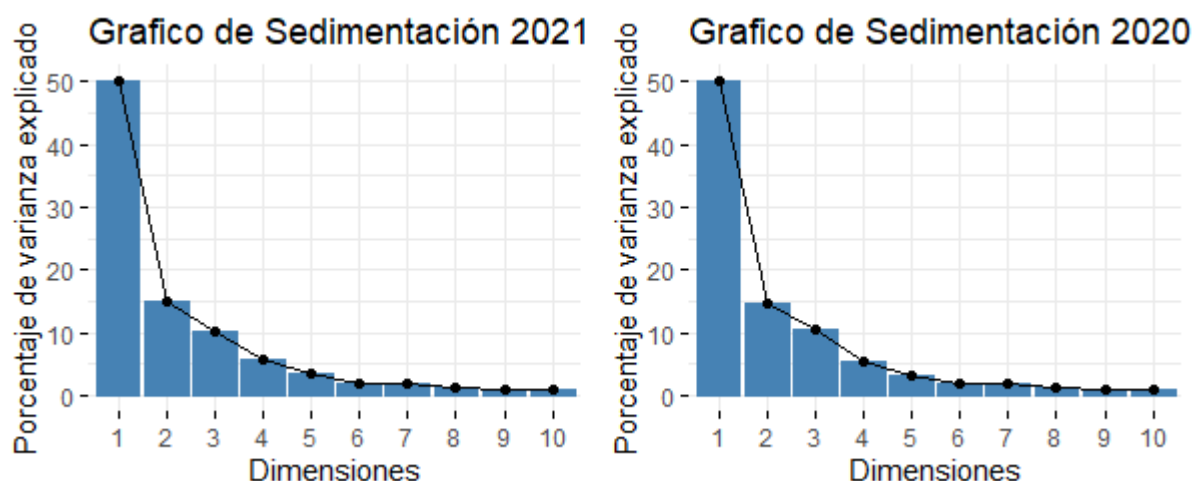


Figura 3: Grafica de varianza por componente años 2021 - 2020, Fuente: Elaboración propia.

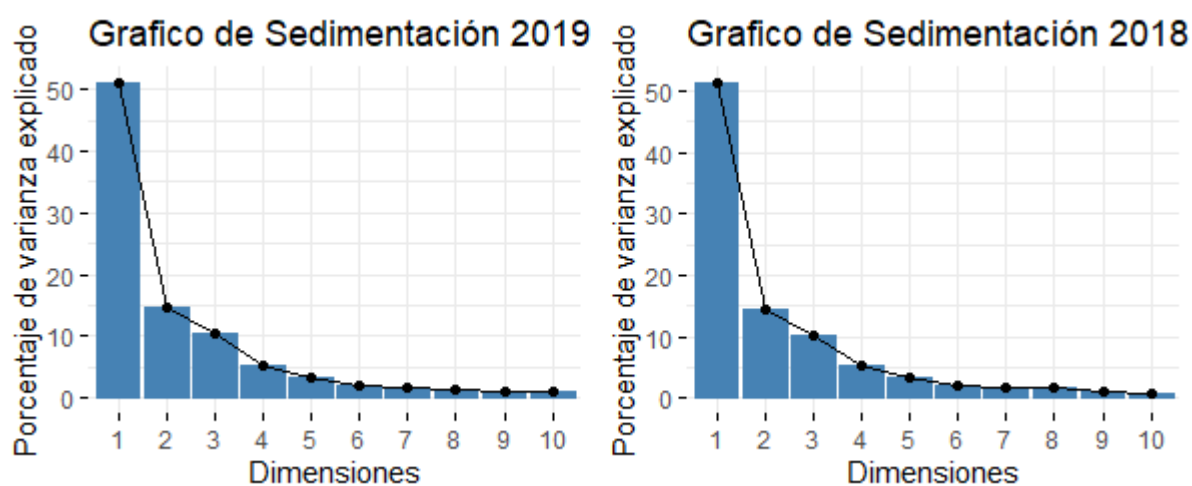


Figura 4: Graficas de varianza por componente años 2019 - 2018, Fuente: Elaboración propia.

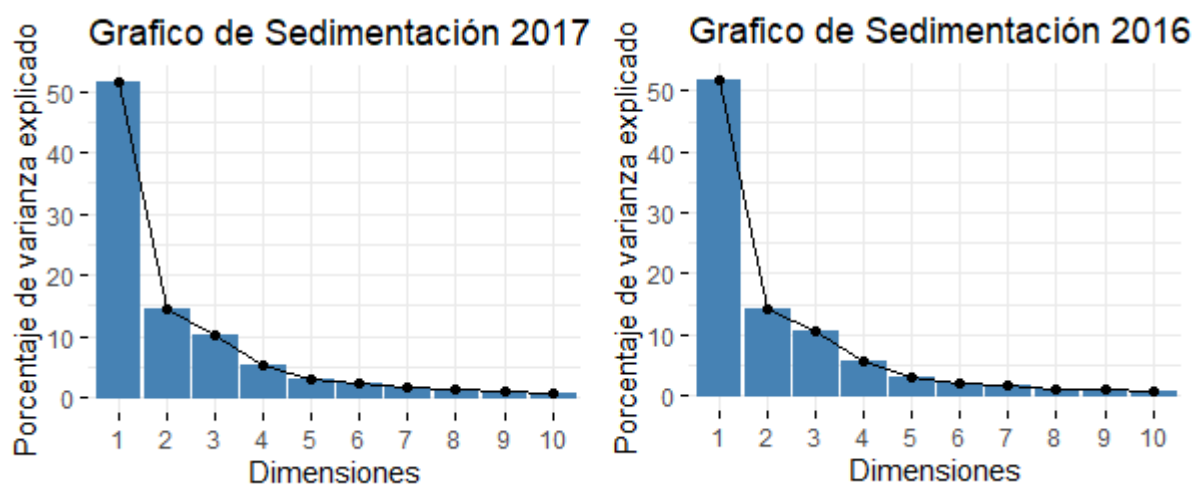


Figura 5: Graficas de varianza por componente años 2017 - 2016, Fuente: Elaboración propia.

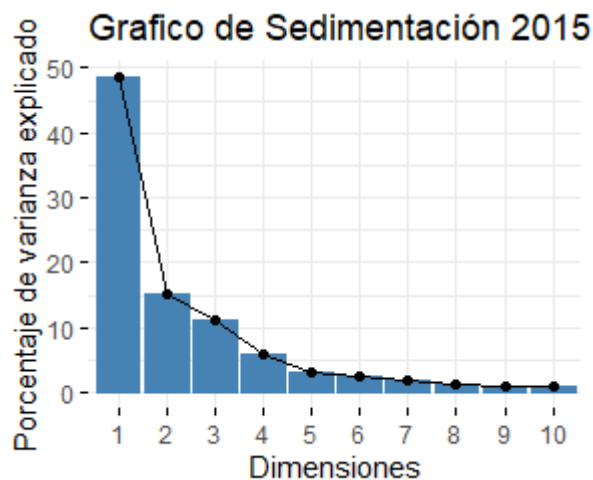


Figura 6: Graficas de varianza por componente año 2015, Fuente: Elaboración propia.

Descripción del plano:

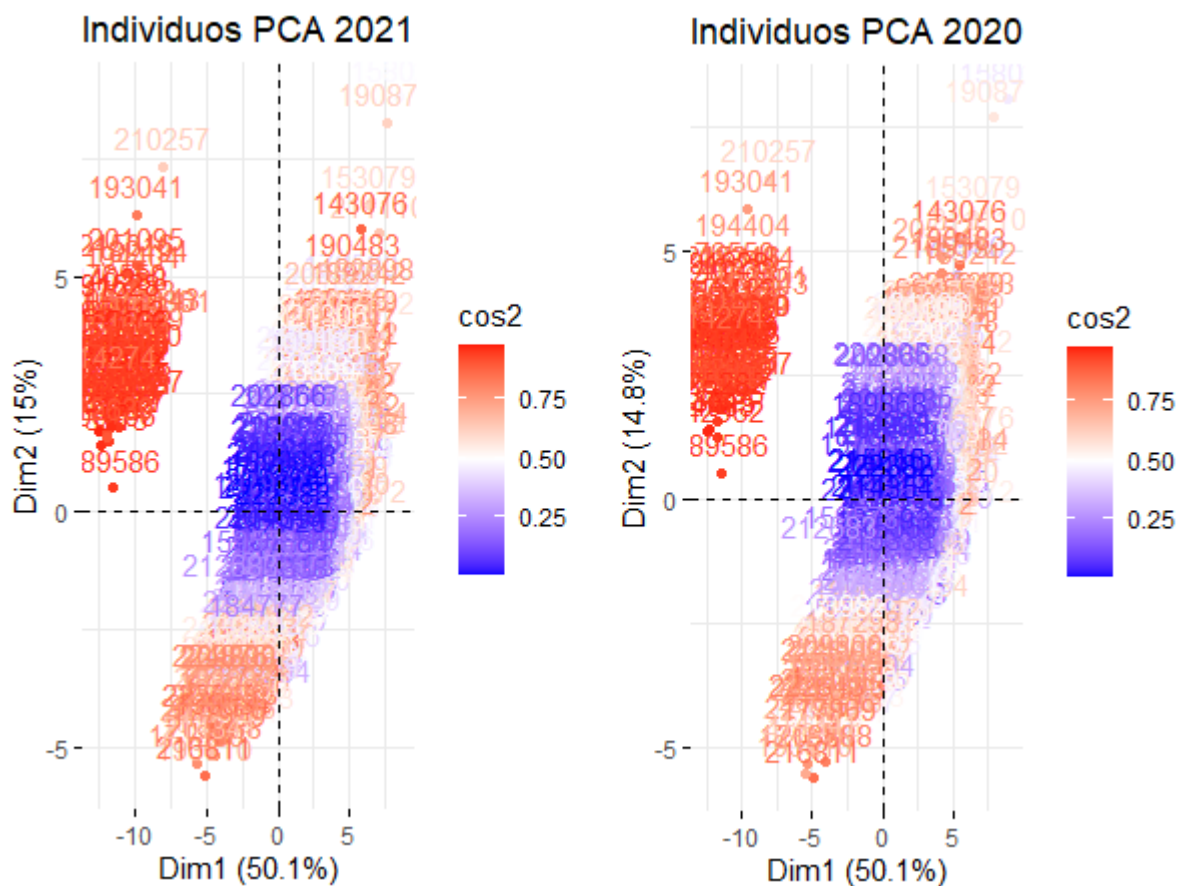


Figura 7: Individuos PCA años 2021-2020, Fuente: Elaboración propia.

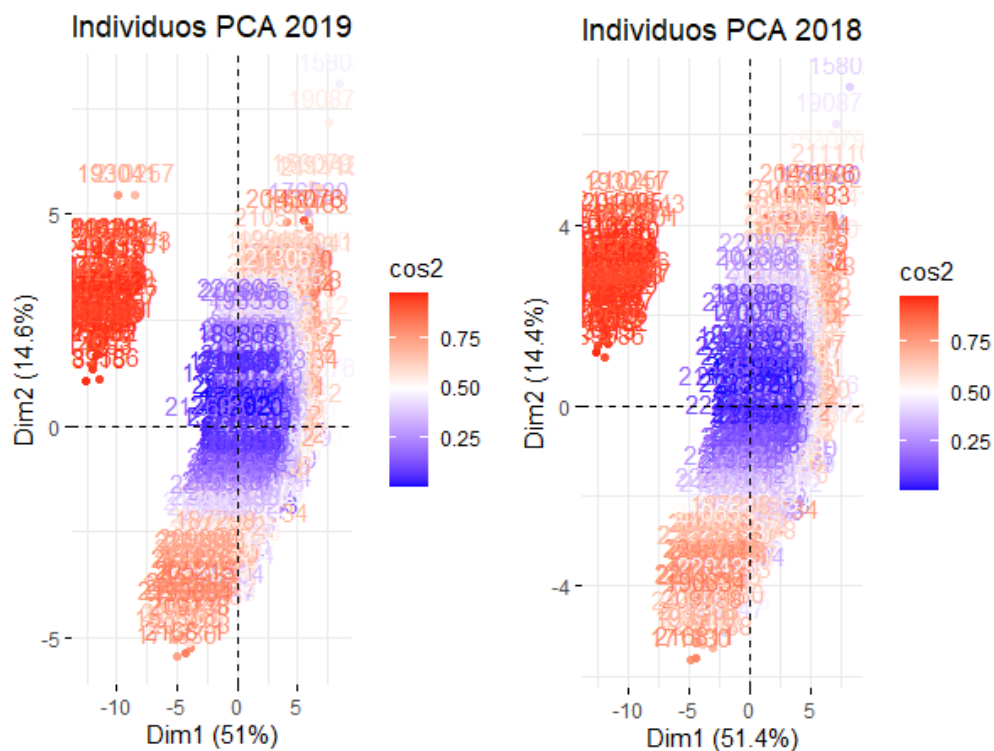
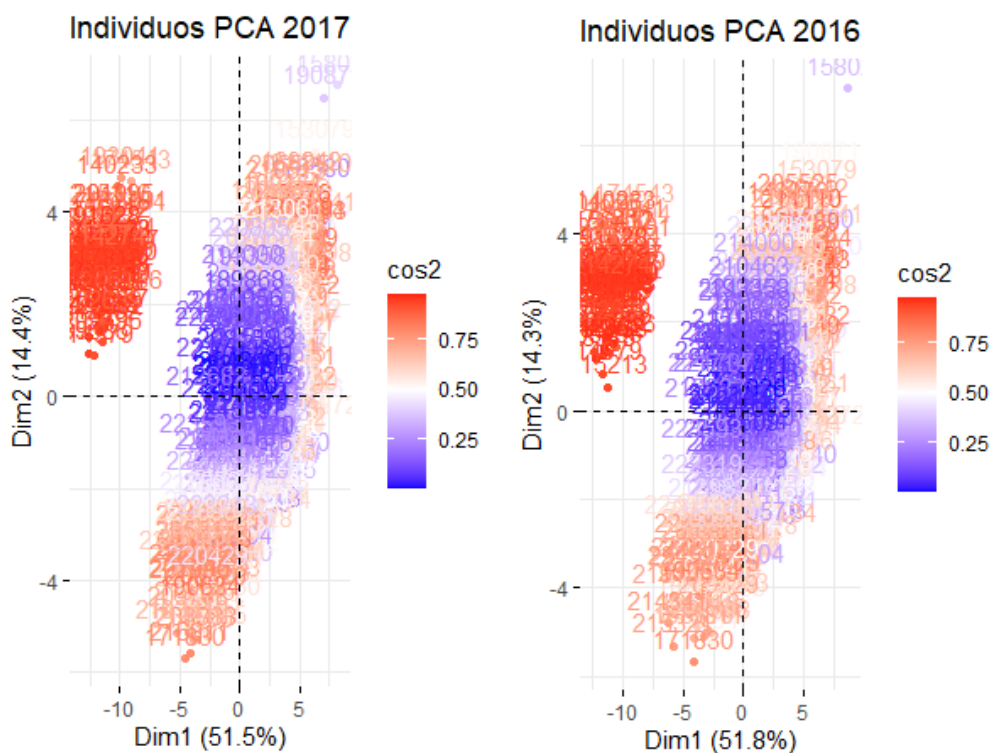


Figura 8: Individuos PCA años 2019-2018, Fuente: Elaboración propia.





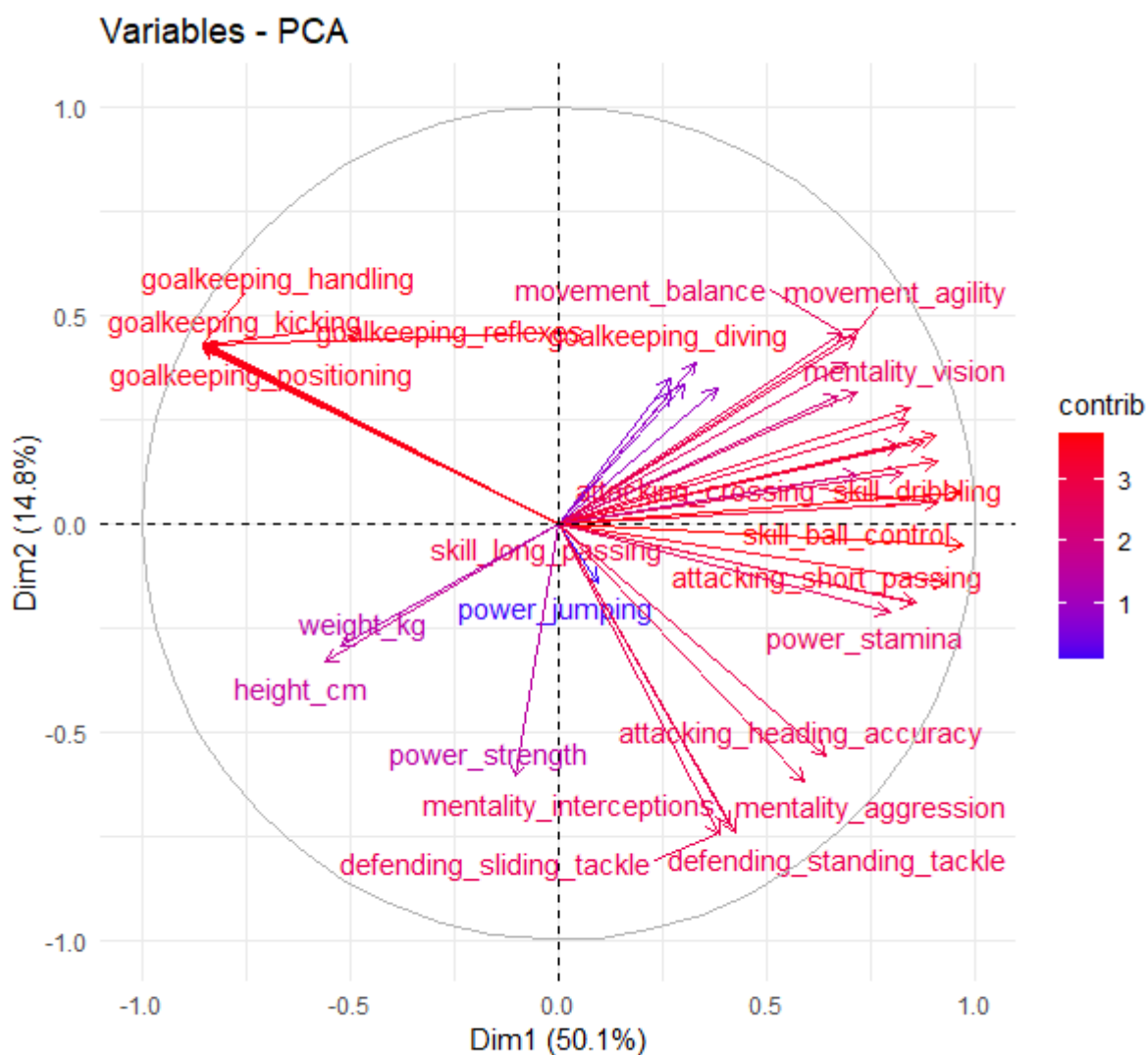


Figura 12: Variables PCA año 2020, Fuente: Elaboración propia.

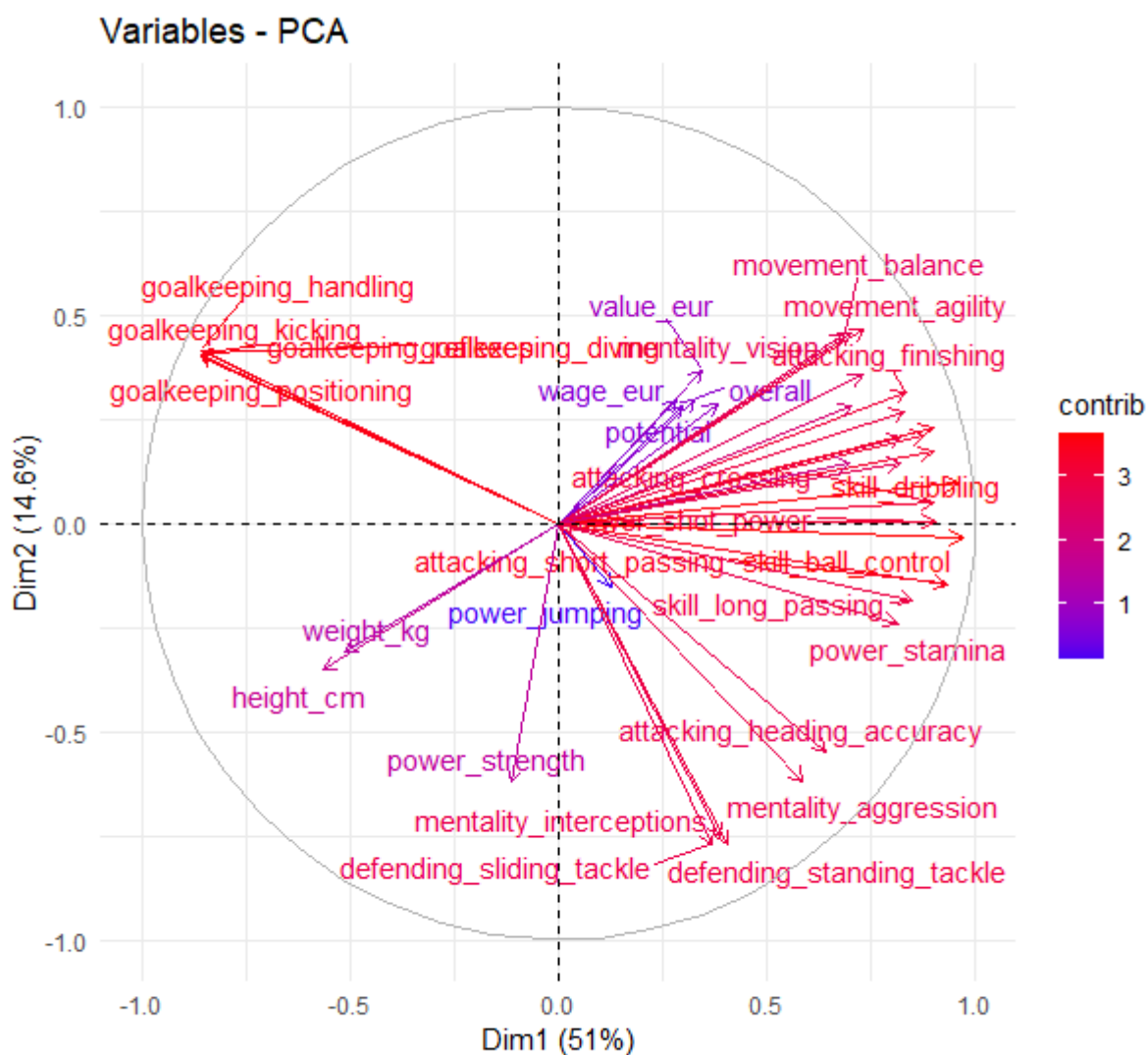


Figura 13: Variables PCA año 2019, Fuente: Elaboración propia.

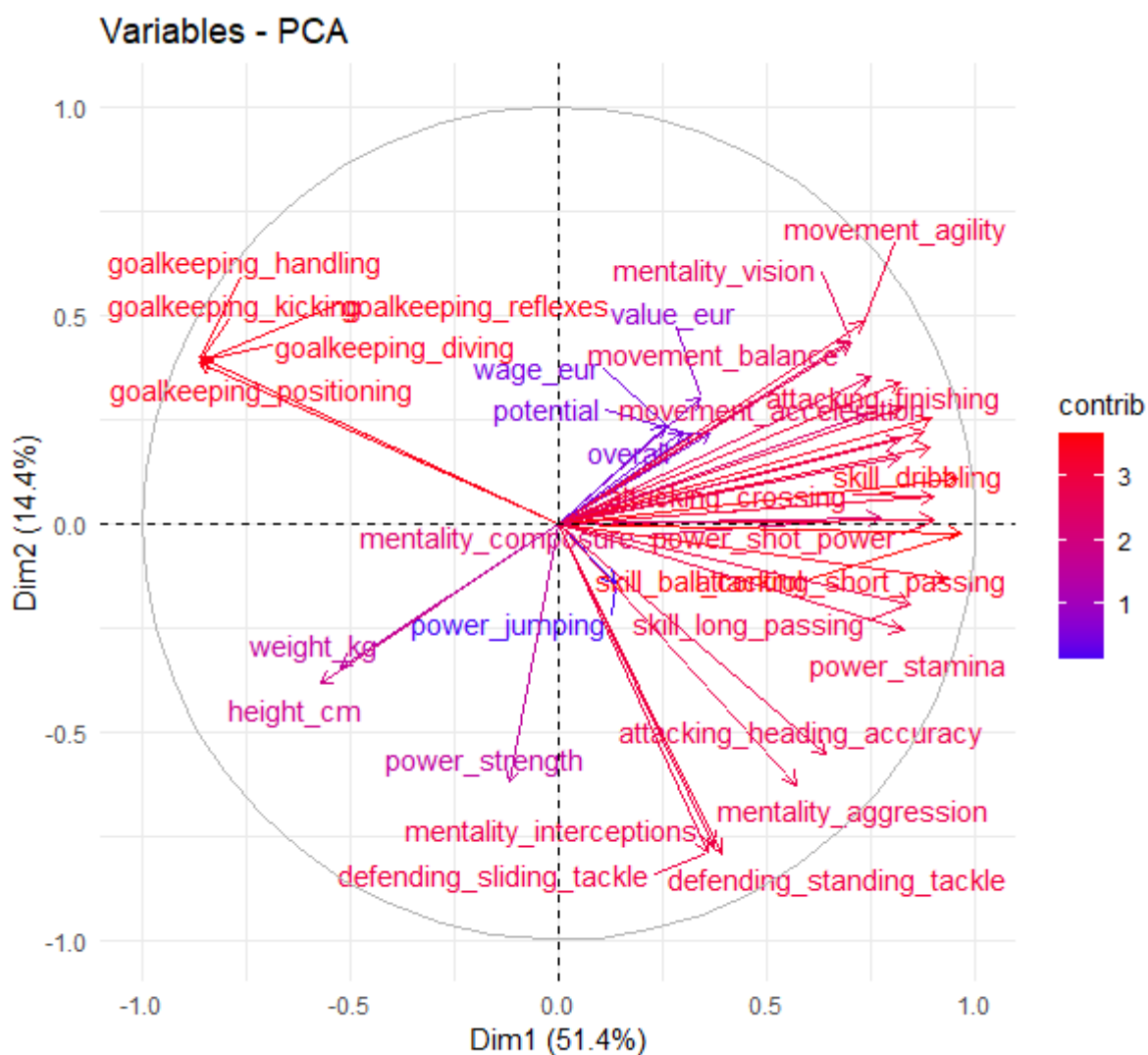


Figura 14: Variables PCA año 2018, Fuente: Elaboración propia.

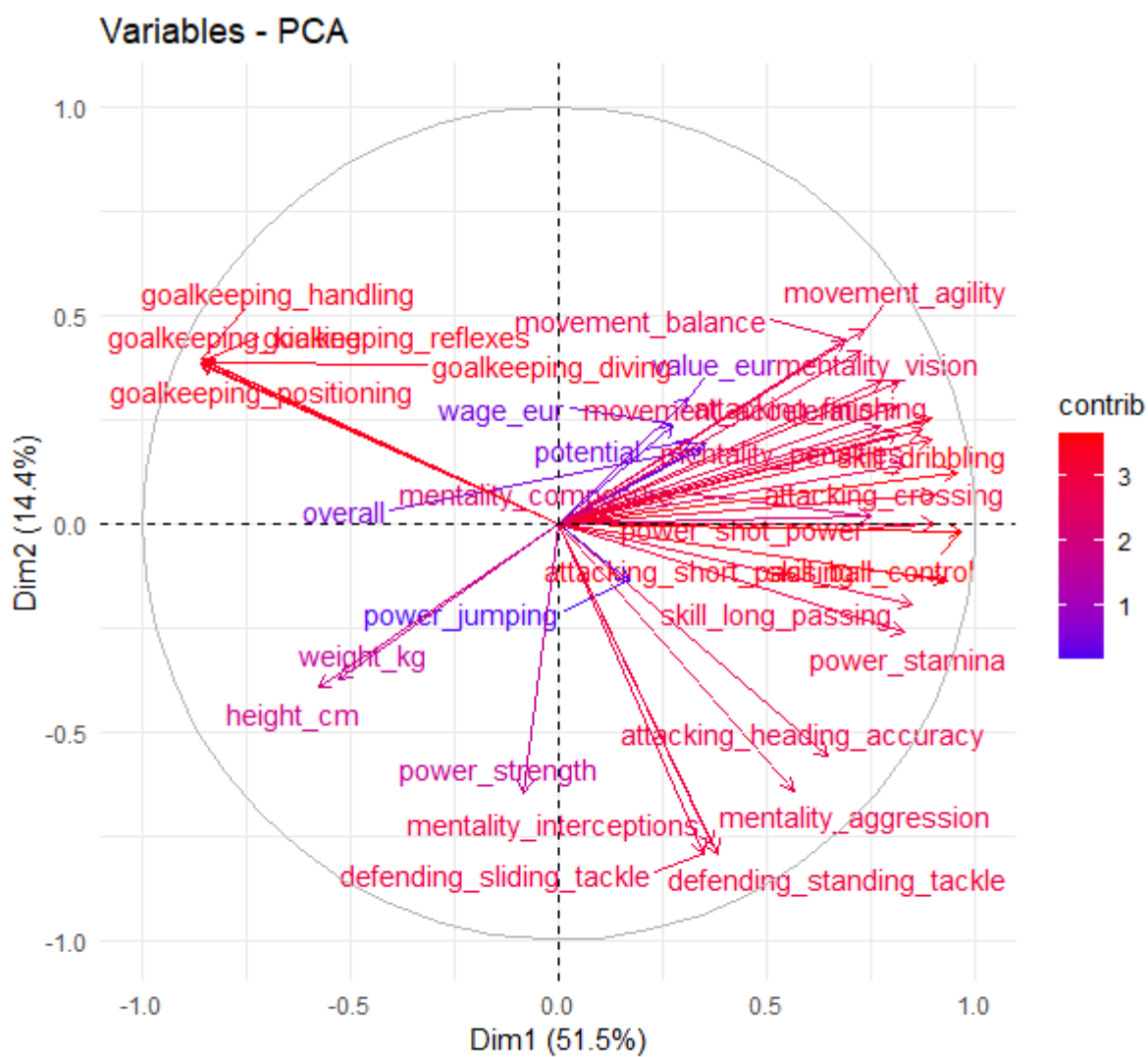


Figura 15: Variables PCA año 2017, Fuente: Elaboración propia.

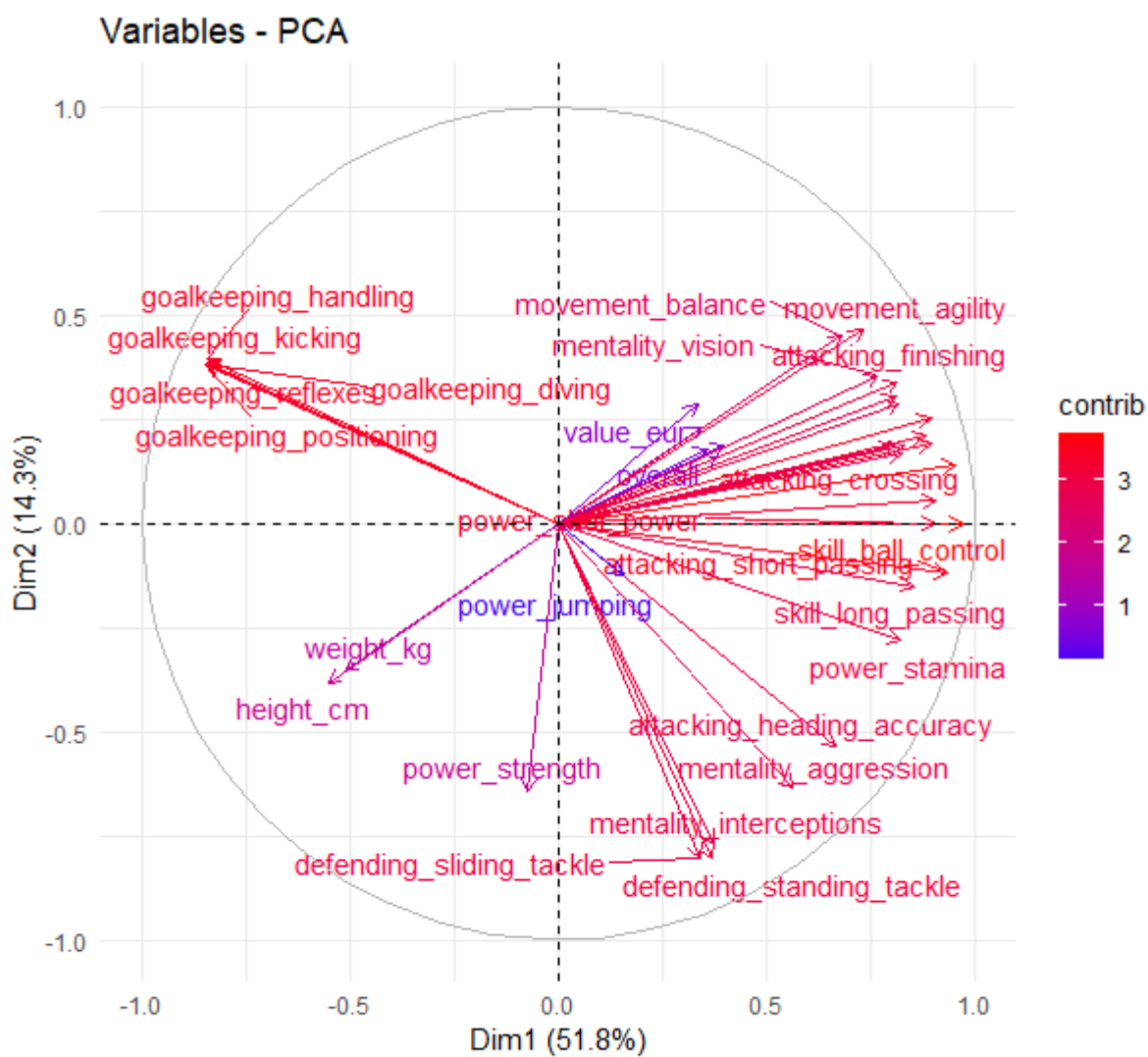


Figura 16: Variables PCA año 2016, Fuente: Elaboración propia.

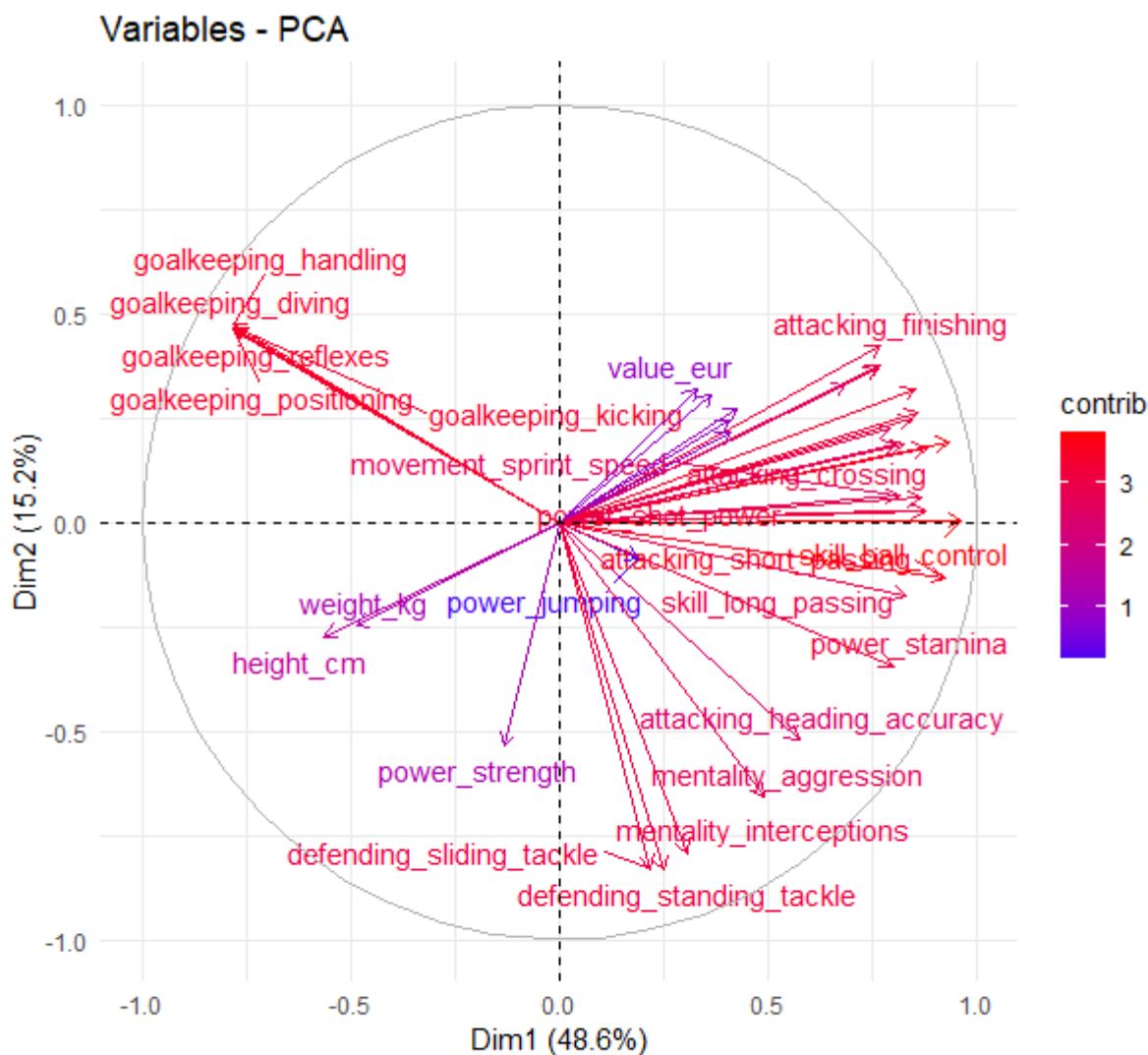


Figura 17: Variables PCA año 2015, Fuente: Elaboración propia.

Para el año 2021 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de regate, dominio de balón, buen posicionamiento en la cancha, fuerza-precisión al momento de pegarle al balón, buena definición, tienen buena resistencia, inteligentes y veloces dentro del campo de juego, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2020 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de regate, dominio de balón, buen posicionamiento en la cancha, fuerza-precisión al momento de pegarle al balón, buena definición, tienen buena resistencia, inteligentes y veloces dentro del campo de juego, se con-

trapone a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2019 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de regate, dominio de balón, buen posicionamiento en la cancha, fuerza-precisión al momento de pegarle al balón, buena definición, tienen buena resistencia, inteligentes y veloces dentro del campo de juego, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2018 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de regate, dominio de balón, buen posicionamiento en la cancha, fuerza-precisión al momento de pegarle al balón, buena definición, tienen buena resistencia, inteligentes y veloces dentro del campo de juego, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2017 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de regate, dominio de balón, buen posicionamiento en la cancha, fuerza-precisión al momento de pegarle al balón, buena definición, tienen buena resistencia, inteligentes y veloces dentro del campo de juego, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2016 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de dominio de balón, regate, pases en distancias reducidas, cruce de balón y fuerza de disparo, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

Para el año 2015 la dimensión 1 está caracterizado por los jugadores que tienen grandes capacidades de dominio de balón, regate, pases en distancias reducidas, inteligencia de juego y fuerza de disparo, se contraponen a jugadores que son altos, de gran peso y que tienen buenas habilidades en la posición de arquero.

A continuación se presentan las cuatro variables representativas y no representativas en la dimensión 1:

Coordenada 2021		Coordenada 2020		Coordenada 2019	
Variable	Dim,1	Variable	Dim,1	Variable	Dim,1
skill ball control	0.9696	skill ball control	0.9717	skill ball control	0.9694
skill dribbling	0.9658	skill dribbling	0.9643	skill dribbling	0.9591
attacking short passing	0.9306	attacking short passing	0.9337	attacking short passing	0.9315
attacking crossing	0.9141	attacking crossing	0.9099	power shot power	0.9035
goalkeeping positioning	-0.8520	goalkeeping diving	-0.8533	goalkeeping handling	-0.8597
goalkeeping diving	-0.8522	goalkeeping positioning	-0.8548	goalkeeping diving	-0.86
goalkeeping handling	-0.8525	goalkeeping reflexes	-0.8549	goalkeeping positioning	-0.8608
goalkeeping reflexes	-0.8532	goalkeeping handling	-0.8550	goalkeeping reflexes	-0.8617

Coordenada 2018		Coordenada 2017		Coordenada 2016	
Variable	Dim,1	Variable	Dim,1	Variable	Dim,1
skill ball control	0.9677	skill ball control	0.9678	skill ball control	0.9702
skill dribbling	0.9551	skill dribbling	0.9556	skill dribbling	0.9531
attacking short passing	0.9322	attacking short passing	0.934	attacking short passing	0.9339
attacking crossing	0.9014	attacking crossing	0.9052	attacking crossing	0.905
goalkeeping handling	-0.8624	goalkeeping handling	-0.8573	goalkeeping handling	-0.8454
goalkeeping positioning	-0.863	goalkeeping diving	-0.8588	goalkeeping positioning	-0.8463
goalkeeping diving	-0.8641	goalkeeping reflexes	-0.8602	goalkeeping diving	-0.8469
goalkeeping reflexes	-0.8657	goalkeeping positioning	-0.8603	goalkeeping reflexes	-0.8496

Coordenada 2015	
Variable	Dim,1
skill ball control	0.9628
skill dribbling	0.9329
attacking short passing	0.9223
mentality vision	0.8793
goalkeeping handling	-0.7818
goalkeeping diving	-0.7832
goalkeeping positioning	-0.7833
goalkeeping reflexes	-0.7853

Tabla 2: Variables más representativas del PCA dimensión 1 años 2021 - 2015, Fuente: Elaboración propia.

Para el año 2021 la dimensión 2 esta caracterizado por los jugadores que tienen gran potencial según el juego FIFA y tienen un valor alto de transferencia, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2020 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones y son muy fuertes, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2019 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones y son muy fuertes, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2018 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones y son muy fuertes, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2017 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones y son muy fuertes, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2016 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones y son muy fuertes, se contraponen de jugadores que tienen buenas habilidades defensivas.

Para el año 2015 la dimensión 2 esta caracterizado por los jugadores que tienen buenas reacciones, se contraponen de jugadores que tienen buenas habilidades defensivas.

Se presentan las cuatro variables representativas y no representativas en la dimensión 2:

Coordenada 2021		Coordenada 2020		Coordenada 2019	
Variable	Dim,2	Variable	Dim,2	Variable	Dim,2
overall	0.4771	movement agility	0.4664	movement agility	0.4653
value eur	0.4656	movement balance	0.456	movement balance	0.4538
potential	0.4565	mentality vision	0.4456	mentality vision	0.4533
mentality vision	0.4518	goalkeeping handling	0.4325	goalkeeping handling	0.4149
mentality aggression	-0.5816	mentality aggression	-0.6199	power strength	-0.622
mentality interceptions	-0.6804	mentality interceptions	-0.7244	mentality interceptions	-0.7539
defending standing tackle	-0.6981	defending standing tackle	-0.7442	defending sliding tackle	-0.7701
defending sliding tackle	-0.7095	defending sliding tackle	-0.7466	defending standing tackle	-0.7721

Coordenada 2018		Coordenada 2017		Coordenada 2016	
Variable	Dim,2	Variable	Dim,2	Variable	Dim,2
movement agility	0.4817	movement agility	0.4661	movement agility	0.4632
mentality vision	0.438	movement balance	0.439	movement balance	0.4492
movement balance	0.4305	mentality vision	0.4142	goalkeeping handling	0.3926
goalkeeping handling	0.3969	goalkeeping handling	0.3929	goalkeeping kicking	0.3854
mentality aggression	-0.6274	power strength	-0.6462	power strength	-0.6422
mentality interceptions	-0.7742	mentality interceptions	-0.7722	mentality interceptions	-0.7712
defending sliding tackle	-0.792	defending sliding tackle	-0.793	defending sliding tackle	-0.8029
defending standing tackle	-0.7925	defending standing tackle	-0.794	defending standing tackle	-0.804

Coordenada 2015	
Variable	Dim,2
goalkeeping handling	0.4746
goalkeeping diving	0.4664
goalkeeping kicking	0.466
goalkeeping reflexes	0.4625
mentality aggression	-0.6585
mentality interceptions	-0.7936
defending sliding tackle	-0.8312
defending standing tackle	-0.8331

Tabla 3: Variables más representativas del PCA dimensión 2 años 2021 - 2015, Fuente: Elaboración propia.

Para los siete años de estudio la dimensión 3 esta caracterizado por los jugadores que no son veloces.

Se presentan las cuarto variables representativas y no representativas en la dimensión 3:

Coordenada 2021		Coordenada 2020		Coordenada 2019	
Variable	Dim,3	Variable	Dim,3	Variable	Dim,3
overall	0.7683	overall	0.8335	overall	0.833
potential	0.7514	potential	0.7856	potential	0.7574
movement reactions	0.6729	value eur	0.7222	value eur	0.7241
value eur	0.6692	movement reactions	0.6988	movement reactions	0.7134
movement sprint speed	-0.2212	movement sprint speed	-0.2463	movement sprint speed	-0.2412
movement acceleration	-0.2413	movement acceleration	-0.2515	movement agility	-0.2416
movement balance	-0.254	movement balance	-0.2539	movement acceleration	-0.2498
movement agility	-0.2722	movement agility	-0.2555	movement balance	-0.2529

Coordenada 2018		Coordenada 2017		Coordenada 2016	
Variable	Dim,3	Variable	Dim,3	Variable	Dim,3
overall	0.8537	overall	0.8576	overall	0.8559
potential	0.751	value eur	0.7631	wage eur	0.8051
value eur	0.7488	movement reactions	0.7402	movement reactions	0.7649
movement reactions	0.7207	potential	0.7243	potential	0.7183
movement agility	-0.1874	movement agility	-0.1467	movement agility	-0.0976
movement sprint speed	-0.2137	movement sprint speed	-0.1845	movement sprint speed	-0.1199
movement acceleration	-0.2322	movement acceleration	-0.19	movement acceleration	-0.1287
movement balance	-0.2588	movement balance	-0.2574	movement balance	-0.2437

Coordenada 2015	
Variable	Dim,3
overall	0.8182
wage eur	0.7344
movement reactions	0.7117
potential	0.6835
movement sprint speed	-0.1301
movement acceleration	-0.1548
movement agility	-0.1576
movement balance	-0.3103

Tabla 4: Variables más representativas del PCA dimensión 3 años 2021 - 2015, Fuente: Elaboración propia.

Los jugadores más representativos por la dimensión 1 son los siguientes:

Coordenada 2021		Coordenada 2020		Coordenada 2019		Coordenada 2018	
Nombre	Dim,1	Nombre	Dim,1	Nombre	Dim,1	Nombre	Dim,1
L. Messi	8.4553	L. Messi	8.8656	L. Messi	8.4577	L. Messi	8.1586
Neymar Jr	7.65	Neymar Jr	7.8057	Neymar Jr	7.5607	Neymar Jr	7.0965
P. Dybala	7.0317	P. Dybala	6.873	P. Dybala	6.8921	A. Sanchez	6.7809
S. Aguero	6.3336	S. Aguero	6.4549	Coutinho	6.6865	P. Dybala	6.7173
R. Firmino	6.2605	Coutinho	6.1066	Marcelo	6.4193	A. Vidal	6.4019
A. Di Maria	6.2243	Marcelo	6.0658	A. Sanchez	6.3301	S. Aguero	6.2242
M. Acuña	5.8168	A. Di Maria	6.0136	S. Aguero	6.2369	Coutinho	5.902
A. Gomez	5.8128	A. Vidal	5.9861	J. Rodriguez	6.0231	L. Suarez	5.8133
A. Vidal	5.6504	R. Firmino	5.941	R. Firmino	5.9076	J. Rodriguez	5.685
Alex Telles	5.6229	M. Acuña	5.5775	Douglas Costa	5.902	Marcelo	5.5884

Coordenada 2017		Coordenada 2016		Coordenada 2015	
Nombre	Dim,1	Nombre	Dim,1	Nombre	Dim,1
L. Messi	8.1569	L. Messi	8.6244	L. Messi	10.189
Neymar Jr	7.0255	A. Vidal	6.754	A. Di Maria	7.595
A. Vidal	6.6605	A. Sanchez	6.5859	A. Vidal	7.1799
A. Sanchez	6.4048	J. Rodriguez	6.4901	S. Aguero	7.07
S. Aguero	6.2774	Neymar Jr	6.3451	Neymar Jr	6.9872
P. Dybala	6.0305	A. Di Maria	6.1012	J. Rodriguez	6.7738
A. Di Maria	5.9975	S. Aguero	6.099	L. Suarez	6.6377
J. Rodriguez	5.9449	L. Suarez	5.7483	A. Sanchez	6.5654
L. Suarez	5.9205	N. Gaitan	5.6758	J. Cuadrado	6.5132
Douglas Costa	5.6718	J. Cuadrado	5.5922	Falcao	6.4965

Tabla 5: Jugadores más representativos del PCA dimensión 1 años 2021 – 2015, Fuente: Elaboración propia.

Los jugadores más representativos por la dimensión 2 son los siguientes:

Coordenada 2021		Coordenada 2020		Coordenada 2019		Coordenada 2018	
Nombre	Dim,2	Nombre	Dim,2	Nombre	Dim,2	Nombre	Dim,2
L. Messi	8.808	L. Messi	8.0719	L. Messi	8.0952	L. Messi	7.0522
Neymar Jr	8.2773	Neymar Jr	7.6946	Neymar Jr	7.191	Neymar Jr	6.2168
Ederson	7.3316	Ederson	6.4685	S. Aguero	5.6793	S. Aguero	5.3887
S. Aguero	6.5242	K. Navas	5.8622	Coutinho	5.6398	P. Dybala	5.0992
K. Navas	6.307	S. Aguero	5.7406	P. Dybala	5.616	Bernard	4.6557
A. Gomez	5.9867	A. Gomez	5.2817	K. Navas	5.4601	A. Gomez	4.6357
P. Dybala	5.9087	P. Dybala	5.1155	Ederson	5.4459	L. Suarez	4.6149
A. Marchesin	5.2356	Neto	4.9768	L. Suarez	5.0159	Coutinho	4.527
G. Rulli	5.0875	Bernard	4.8966	A. Gomez	4.8458	Ederson	4.4994
Douglas Costa	5.0781	Douglas Costa	4.7239	Bernard	4.8324	K. Navas	4.4612

Coordenada 2017		Coordenada 2016		Coordenada 2015	
Nombre	Dim,2	Nombre	Dim,2	Nombre	Dim,2
L. Messi	6.7577	L. Messi	7.2888	L. Messi	8.5596
Neymar Jr	6.4833	Neymar Jr	5.3589	S. Aguero	5.5112
S. Aguero	5.2978	S. Aguero	4.9398	F. Muslera	5.4134
K. Navas	4.7433	Bernard	4.6537	Neymar Jr	5.1462
C. Bravo	4.6641	Coutinho	4.5256	C. Bravo	4.9371
Coutinho	4.606	A. Gomez	4.4593	G. Ochoa	4.8897
P. Dybala	4.5964	C. Bravo	4.3973	L. Suarez	4.8402
G. Ochoa	4.5582	M. Morales	4.3686	Bernard	4.7714
Bernard	4.5337	J. Quintero	4.2949	K. Navas	4.5751
L. Suarez	4.4833	P. Dybala	4.2947	R. Palacio	4.5225

Tabla 6: Jugadores más representativos del PCA dimensión 2 años 2021 – 2015, Fuente: Elaboración propia.

Los jugadores más representativos por la dimensión 3 son los siguientes:

Coordenada 2021		Coordenada 2020		Coordenada 2019		Coordenada 2018	
Nombre	Dim,3	Nombre	Dim,3	Nombre	Dim,3	Nombre	Dim,3
Casemiro	9.3168	L. Messi	9.8767	L. Messi	9.6392	L. Suarez	10.935
L. Messi	8.5673	L. Suarez	7.9721	L. Suarez	9.2463	L. Messi	10.068
Ederson	7.4218	Casemiro	7.6958	Casemiro	7.943	Neymar Jr	7.8854
Fabinho	7.0481	Ederson	7.3592	Neymar Jr	7.4912	G. Higuain	7.8188
L. Suarez	6.4787	Neymar Jr	7.2679	D. Godin	6.4678	S. Aguero	6.4705
Neymar Jr	6.2182	S. Aguero	6.1673	G. Higuain	6.232	Casemiro	6.1803
Marquinhos	6.2049	D. Godin	5.9764	Ederson	5.9319	P. Dybala	5.6805
J. Gimenez	5.8605	Marquinhos	5.9482	P. Dybala	5.7919	D. Godin	5.6523
S. Aguero	5.7933	Fabinho	5.7669	S. Aguero	5.765	A. Sanchez	5.5746
R. Firmino	5.7228	K. Navas	5.6312	Marcelo	5.6428	Thiago Silva	5.4822

Coordenada 2017		Coordenada 2016		Coordenada 2015	
Nombre	Dim,3	Nombre	Dim,3	Nombre	Dim,3
L. Suarez	10.908	L. Messi	11.104	L. Messi	10.356
L. Messi	9.8302	L. Suarez	8.953	L. Suarez	8.4067
Neymar Jr	9.0899	J. Rodriguez	6.5518	Thiago Silva	6.4492
S. Aguero	6.4919	Thiago Silva	6.2867	Falcao	6.3392
G. Higuain	6.4758	Neymar Jr	5.8977	D. Godin	5.6264
J. Rodriguez	6.1802	A. Vidal	5.6335	A. Di Maria	5.5901
Thiago Silva	5.9504	C. Bravo	5.1996	J. Rodriguez	5.4428
D. Godin	5.5385	F. Muslera	5.1777	A. Vidal	5.2831
A. Vidal	5.5103	S. Aguero	5.0681	C. Bravo	5.2782
K. Navas	5.1859	J. Pastore	4.9285	F. Fazio	5.1414

Tabla 7: Jugadores más representativos del PCA dimensión 3 años 2021 – 2015, Fuente: Elaboración propia.

Para todos los años se realiza el mapa de componentes principales y se clasifica por variables, a continuación, solo se muestra del año 2020.

- Jugador Bronce: jugadores con un puntaje menor a 64 (La clasificación es de 0 a 100).
- Jugador Plata: jugadores con un puntaje entre 65 y 74 (La clasificación es de 0 a 100).
- Jugador Oro: jugadores con un puntaje superior a 75 (La clasificación es de 0 a 100).

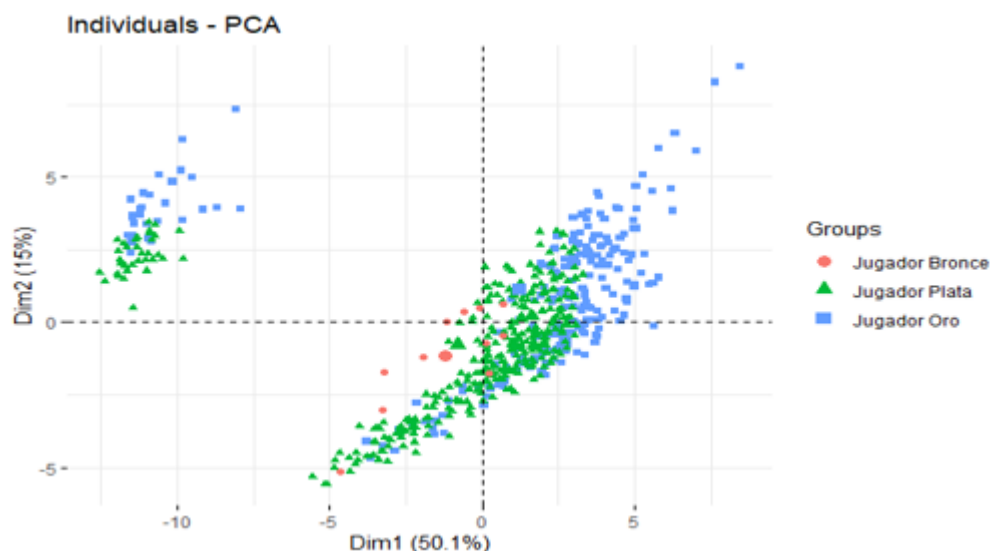


Figura 18: Mapa ACP por puntaje general, Fuente: Elaboración propia.

11.2. Análisis Cluster

Para el análisis cluster se normalizaron las variables implicadas en el análisis de componentes principales por las diferentes escalas de las variables. Como primer paso se calcula la matriz de distancias utilizando el método de distancia euclidiana, el gráfico de las distancias es muy confuso debido a la gran cantidad de variables que tiene la base de datos.

Para poder elegir el número de clústers nos vamos a guiar por tres métodos:

- Método de suma de cuadrados.
- Método Silhouette.
- Método gap stat.

El método de suma de cuadrados al ver los resultados se podría interpretar que el número de clústers a utilizar son dos o tres, pero para verificar el resultado se va a utilizar una función del software Rstudio llamada NbClust el cual compara varios métodos para selección de grupos y muestra cuantos métodos se acomodaron a la cantidad de grupos, los métodos que se utilizan son:

- marriot, trcovw, tracew, friedman, rubin, cindex, db, silhouette, duda, pseudot2, beale, ratkowsky, ball, ptbiserial, gap, frey, mcclain, gamma, gplus, tau, dunn, hubert, sdindex, dindex, y sdbw.

Después de ejecutar la función en Rstudio versión 1.4.1106, nueve métodos proponen que 2 clusters son los mas adecuados para llevar a cabo el informe, ocho métodos proponen que la cantidad de grupos sean tres y tres métodos proponen cuatro grupos.

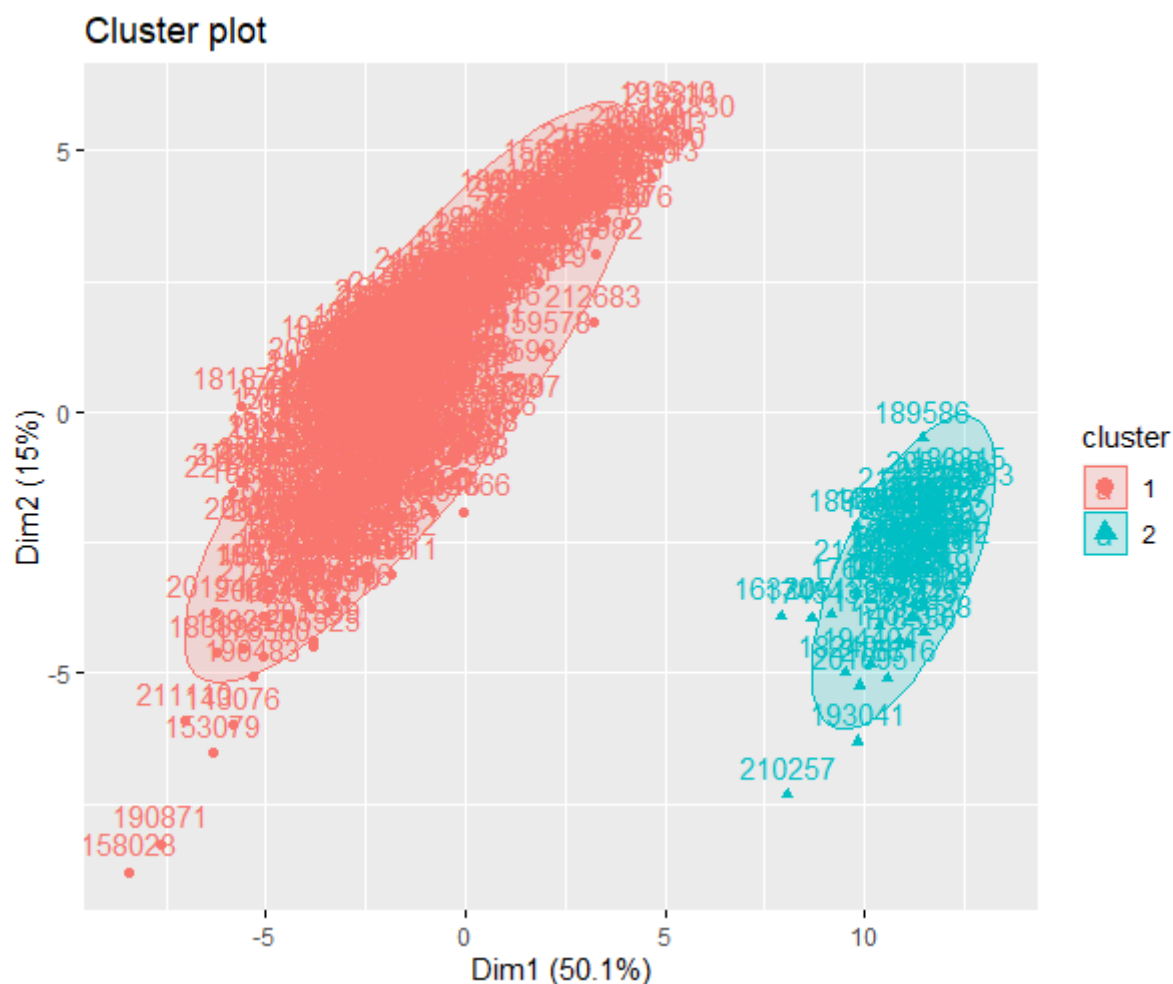


Figura 19: Grafico cluster con 2 grupos, Fuente: Elaboración propia.

Al interpretar el gráfico y después de observar las cualidades de los dos grupos, no es una buena clasificación, ya que el grupo de la izquierda contiene la mayoría de los jugadores que tienen distintas cualidades y el grupo de la derecha pertenece a los jugadores que ocupan la posición de arquero, entonces por ende pasamos a aumentar el número de grupos a tres que es el segundo más óptimo o cuatro grupos, lo mencionado anteriormente aplica para los siete años de estudio ya que la información de los jugadores año tras año no cambia mucho.

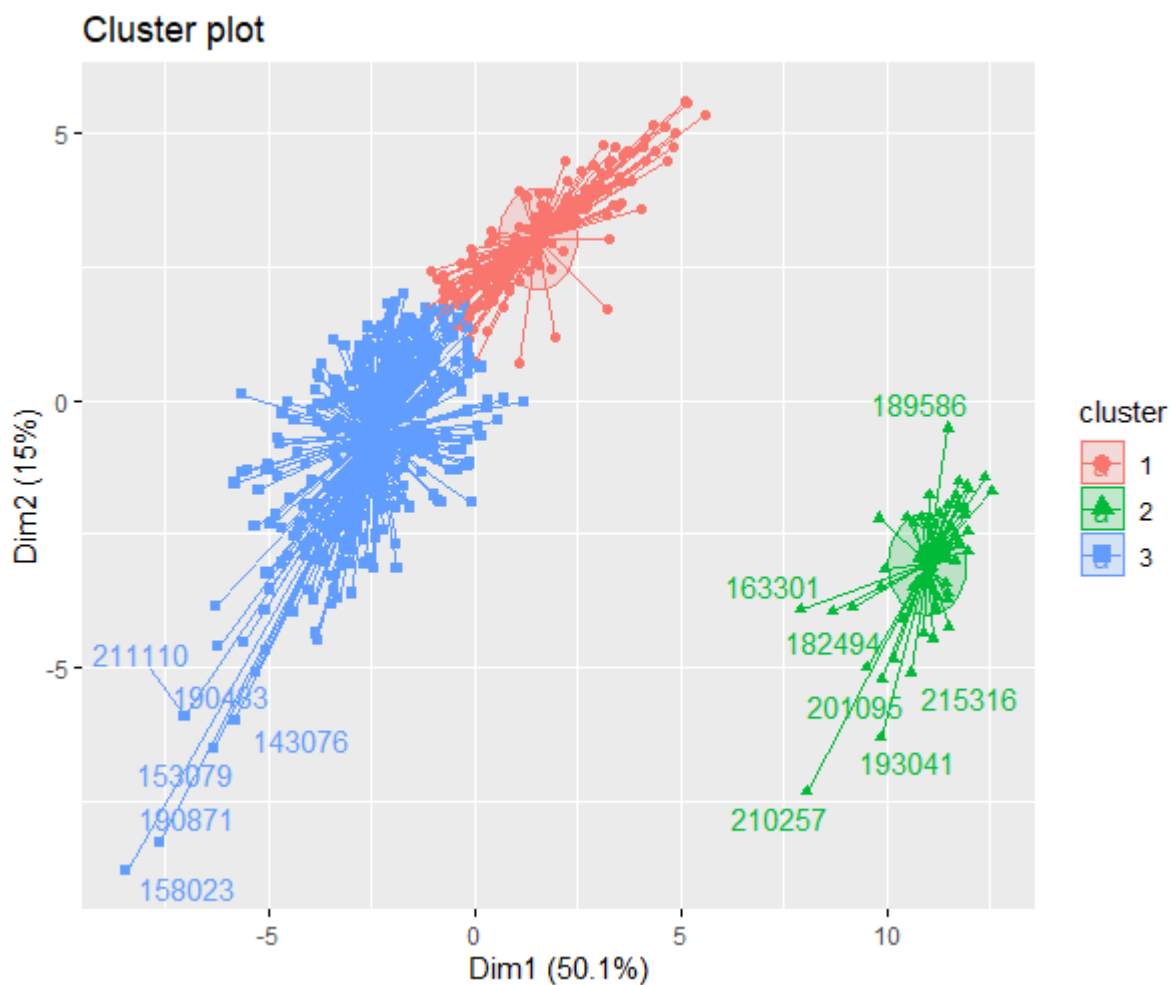


Figura 20: Grafico cluster con 3 grupos, Fuente: Elaboración propia.

Para el año 2021 el grupo 1 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2021 el grupo 2 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2021 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2021			2021		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
power strength	0.8107	Grupo 1	goalkeeping handling	2.828	Grupo 2
defending sliding tackle	0.7234	Grupo 1	goalkeeping diving	2.8275	Grupo 2
defending standing tackle	0.7117	Grupo 1	goalkeeping reflexes	2.8238	Grupo 2
mentality interceptions	0.6987	Grupo 1	goalkeeping positioning	2.8201	Grupo 2
attacking heading accuracy	0.6311	Grupo 1	goalkeeping kicking	2.8124	Grupo 2
mentality aggression	0.5421	Grupo 1	height cm	1.0964	Grupo 2
power jumping	0.1611	Grupo 1	weight kg	0.9996	Grupo 2

2021		
Variable	Grupo 3	Clasificación
mentality positioning	0.5876	Grupo 3
skill curve	0.5817	Grupo 3
power long shots	0.5745	Grupo 3
attacking finishing	0.5647	Grupo 3
skill dribbling	0.5594	Grupo 3
attacking volleys	0.5581	Grupo 3
movement agility	0.5429	Grupo 3

Tabla 8: Clasificación jugadores a los 3 grupos año 2021, Fuente: Elaboración propia.

Para el año 2020 el grupo 1 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2020 el grupo 2 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Para el año 2020 el grupo 3 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Se muestran algunas variables que tomo cada grupo:

2020			2020		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
defending standing tackle	0.7656	Grupo 1	mentality positioning	0.6110	Grupo 2
defending sliding tackle	0.7646	Grupo 1	attacking finishing	0.6015	Grupo 2
power strength	0.7633	Grupo 1	power long shots	0.6013	Grupo 2
mentality interceptions	0.7276	Grupo 1	skill curve	0.5942	Grupo 2
attacking heading accuracy	0.6348	Grupo 1	attacking volleys	0.5804	Grupo 2
mentality aggression	0.5658	Grupo 1	skill dribbling	0.5762	Grupo 2
power jumping	0.1854	Grupo 1	mentality vision	0.5634	Grupo 2

2020		
Variable	Grupo 3	Clasificación
goalkeeping handling	2.8283	Grupo 3
goalkeeping diving	2.8245	Grupo 3
goalkeeping reflexes	2.8237	Grupo 3
goalkeeping positioning	2.8206	Grupo 3
goalkeeping kicking	2.8112	Grupo 3
height cm	1.0811	Grupo 3
weight kg	1.0144	Grupo 3

Tabla 9: Clasificación jugadores a los 3 grupos año 2020, Fuente: Elaboración propia.

Para el año 2019 el grupo 1 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2019 el grupo 2 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2019 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2019			2019		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
goalkeeping handling	2.8259	Grupo 1	defending standing tackle	0.7531	Grupo 2
goalkeeping reflexes	2.8243	Grupo 1	defending sliding tackle	0.7500	Grupo 2
goalkeeping diving	2.8239	Grupo 1	power strength	0.7369	Grupo 2
goalkeeping positioning	2.8187	Grupo 1	mentality interceptions	0.7122	Grupo 2
goalkeeping kicking	2.8074	Grupo 1	attacking heading accuracy	0.6250	Grupo 2
height cm	1.0986	Grupo 1	mentality aggression	0.5262	Grupo 2
weight kg	1.0070	Grupo 1	power jumping	0.1242	Grupo 2

2019		
Variable	Grupo 3	Clasificación
overall	0.1751	Grupo 3
potential	0.1497	Grupo 3
value eur	0.1711	Grupo 3
wage eur	0.1325	Grupo 3
attacking crossing	0.5400	Grupo 3
attacking finishing	0.6022	Grupo 3
attacking short passing	0.4576	Grupo 3

Tabla 10: Clasificación jugadores a los 3 grupos año 2019, Fuente: Elaboración propia.

Para el año 2018 el grupo 1 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2018 el grupo 2 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2018 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2018			2018		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
goalkeeping diving	2.8231	Grupo 1	defending sliding tackle	0.7758	Grupo 2
goalkeeping handling	2.8230	Grupo 1	defending standing tackle	0.7754	Grupo 2
goalkeeping reflexes	2.8223	Grupo 1	power strength	0.7700	Grupo 2
goalkeeping positioning	2.8140	Grupo 1	mentality interceptions	0.7220	Grupo 2
goalkeeping kicking	2.8039	Grupo 1	attacking heading accuracy	0.6511	Grupo 2
height cm	1.1070	Grupo 1	mentality aggression	0.5588	Grupo 2
weight kg	1.0233	Grupo 1	power jumping	0.1099	Grupo 2

2018		
Variable	Grupo 3	Clasificación
mentality positioning	0.6194	Grupo 3
skill curve	0.6083	Grupo 3
power long shots	0.6081	Grupo 3
movement agility	0.5923	Grupo 3
attacking finishing	0.5894	Grupo 3
skill dribbling	0.5790	Grupo 3
attacking volleys	0.5772	Grupo 3

Tabla 11: Clasificación jugadores a los 3 grupos año 2018, Fuente: Elaboración propia.

Para el año 2017 el grupo 1 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2017 el grupo 2 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2017 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2017			2017		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
goalkeeping diving	2.8193	Grupo 1	power strength	0.7737	Grupo 2
goalkeeping reflexes	2.8173	Grupo 1	defending sliding tackle	0.7657	Grupo 2
goalkeeping handling	2.8166	Grupo 1	defending standing tackle	0.7645	Grupo 2
goalkeeping positioning	2.8108	Grupo 1	mentality interceptions	0.7069	Grupo 2
goalkeeping kicking	2.8006	Grupo 1	attacking heading accuracy	0.6524	Grupo 2
height cm	1.1235	Grupo 1	mentality aggression	0.5612	Grupo 2
weight kg	1.013	Grupo 1	power jumping	0.1062	Grupo 2

2017		
Variable	Grupo 3	Clasificación
overall	0.1538	Grupo 3
potential	0.1574	Grupo 3
value eur	0.1419	Grupo 3
wage eur	0.1200	Grupo 3
attacking crossing	0.5441	Grupo 3
attacking finishing	0.5879	Grupo 3
attacking short passing	0.4704	Grupo 3

Tabla 12: Clasificación jugadores a los 3 grupos año 2017, Fuente: Elaboración propia.

Para el año 2016 el grupo 1 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2016 el grupo 2 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2016 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2016			2016		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
defending sliding tackle	0.6933	Grupo 1	goalkeeping diving	2.8199	Grupo 2
defending standing tackle	0.6846	Grupo 1	goalkeeping reflexes	2.8143	Grupo 2
power strength	0.6039	Grupo 1	goalkeeping handling	2.8113	Grupo 2
mentality interceptions	0.6035	Grupo 1	goalkeeping positioning	2.8026	Grupo 2
attacking heading accuracy	0.5190	Grupo 1	goalkeeping kicking	2.7924	Grupo 2
mentality aggression	0.4441	Grupo 1	height cm	1.1045	Grupo 2
power jumping	0.0649	Grupo 1	weight kg	1.0385	Grupo 2

2016		
Variable	Grupo 3	Clasificación
skill curve	0.6585	Grupo 3
mentality positioning	0.6567	Grupo 3
power long shots	0.6562	Grupo 3
attacking finishing	0.6376	Grupo 3
skill dribbling	0.6332	Grupo 3
movement agility	0.6326	Grupo 3
attacking volleys	0.6293	Grupo 3

Tabla 13: Clasificación jugadores a los 3 grupos año 2016, Fuente: Elaboración propia.

Para el año 2015 el grupo 1 clasificó jugadores con buenas habilidades de arquero, con buen potencial, con una gran altura y también son de gran peso.

Para el año 2015 el grupo 2 clasificó a los jugadores con buena puntería de disparo, con buena fuerza de salto y de cuerpo, adicional a ello son jugadores con buenas habilidades defensivas.

Para el año 2015 el grupo 3 clasificó jugadores habilidosos con el balón, con un buen valor en el mercado de transferencias, tienen mucha inteligencia de juego, buen pegue de balón y son muy veloces.

Se muestran algunas variables que tomo cada grupo:

2015			2015		
Variable	Grupo 1	Clasificación	Variable	Grupo 2	Clasificación
goalkeeping diving	2.8192	Grupo 1	defending sliding tackle	0.4980	Grupo 2
goalkeeping reflexes	2.8110	Grupo 1	defending standing tackle	0.4877	Grupo 2
goalkeeping handling	2.8083	Grupo 1	mentality interceptions	0.4388	Grupo 2
goalkeeping positioning	2.7988	Grupo 1	attacking heading accuracy	0.3419	Grupo 2
goalkeeping kicking	2.7906	Grupo 1	power strength	0.3175	Grupo 2
height cm	1.1136	Grupo 1	mentality aggression	0.3130	Grupo 2

2015		
Variable	Grupo 3	Clasificación
power long shots	0.7781	Grupo 3
mentality positioning	0.7678	Grupo 3
skill curve	0.7558	Grupo 3
attacking volleys	0.7555	Grupo 3
attacking finishing	0.7550	Grupo 3
skill dribbling	0.7549	Grupo 3

Tabla 14: Clasificación jugadores a los 3 grupos año 2015, Fuente: Elaboración propia.

Para los siete años de estudio se tiene el mismo comportamiento en los tres grupos, un grupo conformado por atacantes, otro por defensas y un último conformado por arqueros, Ahora se realiza una segunda clasificación cluster por cada grupo mencionado anteriormente para poder identificar los jugadores más destacados de cada grupo, para esta nueva clasificación solo se va a elegir dos cluster por cada grupo.

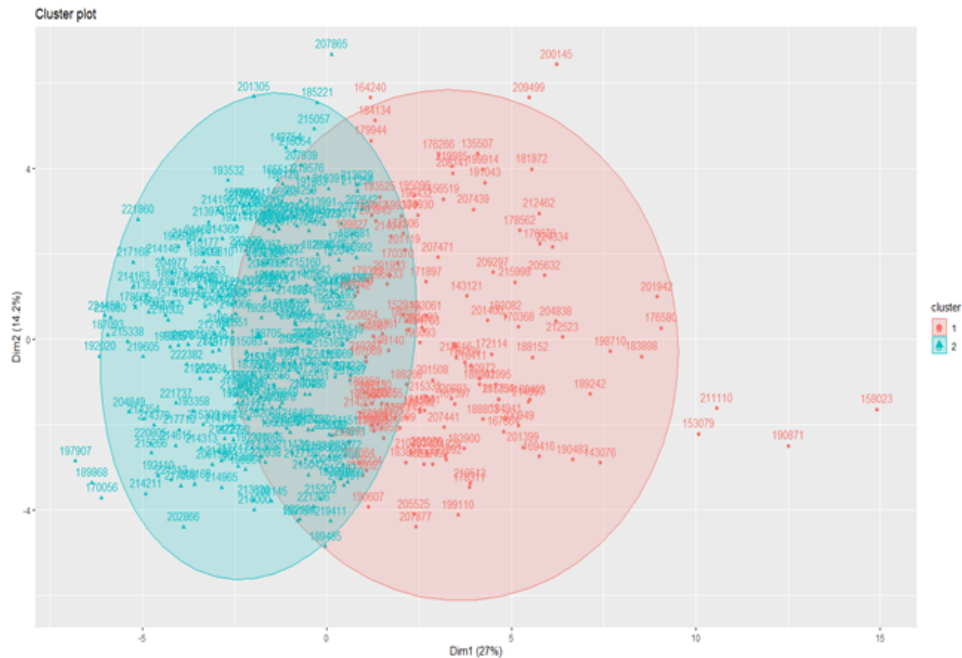


Figura 21: Segunda clasificación cluster con 2 grupos, Fuente: Elaboración propia.

Para el año 2021 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2021 Cluster 1		2021 Cluster 2		2021 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
L. Messi	Marquinhos	G. Ochoa	Ederson	D. Godin	J. Gimenez
Neymar Jr	N. Tagliafico	D. Ospina	K. Navas	Miranda	S. Coates
S. Aguero	G. Paulista	C. Vargas	G. Rulli	W. Barrios	G. Pezzella
Casemiro	S. Arias	G. Arias	F. Muslera	Sidnei	C. Izquierdoz
P. Dybala	Luiz Gustavo	Gabriel	Neto	N. Otamendi	L. Cabrera
Fabinho	Fernando	L. Chichizola	A. Marchesin	M. Musacchio	F. Fazio
L. Suarez	J. Mascherano	A. Blake	W. Benitez	J. Murillo	F. Fernandez
A. Di Maria	F. Fabra	H. Gonzalez	E. Andrada	Hilton	Marcelo
R. Firmino	L. Rodriguez	W. Caballero	J. Musso	H. Moreno	O. Murillo
A. Gomez	C. Ortiz	J. Corona	S. Romero	E. Balanta	L. Lopez

Tabla 15: Segunda clasificación 2021, Fuente: Elaboración propia.

Para el año 2020 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2020 Cluster 1		2020 Cluster 2		2020 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
D. Godin	J. Gimenez	Ederson	G. Ochoa	L. Messi	Marquinhos
Miranda	S. Coates	K. Navas	G. Arias	Neymar Jr	N. Tagliafico
N. Otamendi	Jardel	Neto	C. Vargas	S. Aguero	Fernando
G. Paulista	F. Fazio	G. Rulli	Matheus	L. Suarez	Luiz Gustavo
Sidnei	L. Cabrera	F. Muslera	A. Aguerre	P. Dybala	J. Mascherano
M. Musacchio	G. Pezzella	E. Andrada	A. Blake	Casemiro	J. Lerma
J. Murillo	G. Maripan	W. Benitez	L. Chichizola	Fernandinho	S. Arias
W. Barrios	C. Zapata	S. Romero	Gabriel	Thiago Silva	S. Rondon
J. Pinola	Marcelo	D. Ospina	W. Caballero	R. Firmino	D. Suarez
Hilton	C. Izquierdoz	F. Armani	Rafael	A. Di Maria	Dyego Sousa

Tabla 16: Segunda clasificacion 2020, Fuente: Elaboración propia.

Para el año 2019 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2019 Cluster 1		2019 Cluster 2		2019 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
F. Armani	K. Navas	D. Godin	J. Gimenez	L. Messi	Luiz Gustavo
Matheus	Ederson	N. Otamendi	S. Coates	Neymar Jr	M. Fernandes
D. Ospina	Neto	Miranda	Jardel	L. Suarez	G. Medel
Guilherme	G. Rulli	Marquinhos	F. Fazio	P. Dybala	J. Mascherano
L. Chichizola	F. Muslera	G. Paulista	Marcelo	S. Aguero	M. Rojo
G. Ochoa	S. Romero	J. Murillo	Sidnei	Casemiro	J. Lerma
W. Benitez	C. Bravo	M. Musacchio	G. Pezzella	Coutinho	N. Tagliafico
W. Caballero	N. Guzman	I. Marcone	C. Izquierdoz	J. Rodriguez	R. Battaglian
A. Blake	A. Marchesin	D. Abraham	C. Zapata	Thiago Silva	Fernando
C. Vargas	J. Orozco	W. Barrios	Juan Jesus	G. Higuain	Ismaily

Tabla 17: Segunda clasificacion 2019, Fuente: Elaboración propia.

Para el año 2018 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2018 Cluster 1		2018 Cluster 2		2018 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
F. Armani	K. Navas	D. Godin	J. Gimenez	J. Mascherano	L. Messi
S. Torrico	Ederson	Miranda	F. Fazio	M. Rojo	Neymar
C. Vargas	G. Rulli	Marquinhos	S. Coates	G. Medel	L. Suarez
Rafael	C. Bravo	M. Musacchio	G. Cabral	Fernando	G. Higuain
Dani H	F. Muslera	N. Otamendi	Jardel	N. Tagliafico	S. Aguero
S. Sosa	S. Romero	Sidnei	Eric Botteghin	D. Zapata	A. Sanchez
Gabriel	Neto	H. Moreno	Daniilo	G. Pizarro	P. Dybala
Da Costa	W. Caballero	Luiz Gustavo	Marcelo	F. Gago	Thiago Silva
N. Vikonis	G. Ochoa	J. Murillo	F. Fernandez	Ismaily	Marcelo
J. Cuadrado	D. Ospina	I. Marcone	G. Pezzella	G. Escalante	A. Vidal

Tabla 18: Segunda clasificacion 2018, Fuente: Elaboración propia.

Para el año 2017 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2017 Cluster 1		2017 Cluster 2		2017 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
J. Mascherano	L. Messi	S. Torrico	C. Bravo	D. Godin	J. Gimenez
G. Medel	Neymar	F. Armani	K. Navas	Miranda	Jardel
Mario F	L. Suarez	Rafael	F. Muslera	N. Otamendi	G. Cabral
C. Ansaldi	S. Agüero	N. Vikonis	G. Rulli	M. Musacchio	S. Coates
Fabinho	Thiago Silva	Gabriel	G. Ochoa	Luiz Gustavo	W. Morgan
Fernando	G. Higuain	R. Rey	Neto	Marquinhos	Daniilo
O. Alonso	J. Rodriguez	C. Toselli	D. Ospina	J. Murillo	F. Fernandez
G. Escalante	A. Vidal	A. Dominguez	S. Romero	Sidnei	B. Espinosa
Wendell	A. Di Maria	C. Vargas	Ederson	H. Moreno	Marcelo
S. Cristóforo	A. Sanchez	S. Sosa	A. Marchesin	Lucas Leiva	R. Civelli

Tabla 19: Segunda clasificación 2017, Fuente: Elaboración propia.

Para el año 2016 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2016 Cluster 1		2016 Cluster 2		2016 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
F. Coloccini	D. Godin	Ederson	C. Bravo	J. Mascherano	L. Messi
Lucas Leiva	Miranda	A. Dominguez	F. Muslera	Luiz Gustavo	L. Suarez
Gabriel	N. Otamendi	M. Caranta	G. Ochoa	M. Rojo	Neymar
Hilton	M. Musacchio	Dani H	G. Rulli	G. Medel	Thiago Silva
D. Abraham	Marquinhos	Da Costa	Neto	Mario F	J. Rodriguez
D. Reyes	Jardel	E. Andrada	K. Navas	C. Ansaldi	S. Agüero
Djavan	Juan Jesus	F. Armani	D. Ospina	Fabinho	A. Sanchez
F. Roncaglia	H. Moreno	S. Viera	W. Caballero	Daniilo	A. Vidal
S. Garcia	J. Gimenez	N. Vikonis	Guilherme	Fernando	A. Di Maria
J. Gandolfi	F. Fernandez	W. Benitez	Matheus	Wendell	R. Oscar

Tabla 20: Segunda clasificación 2016, Fuente: Elaboración propia.

Para el año 2015 algunos jugadores los clasificados a los nuevos grupos son los siguientes:

2015 Cluster 1		2015 Cluster 2		2015 Cluster 3	
Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
C. Bravo	Rafael	F. Roncaglia	D. Godin	J. Mascherano	L. Messi
F. Muslera	A. Dominguez	Fabinho	Miranda	Luiz Gustavo	L. Suarez
G. Ochoa	N. F. Armani	Gabriel	N. Otamendi	Fernando	Falcao
K. Navas	Dani H	E. Insua	F. Fazio	M. Rojo	Thiago Silva
W. Caballero	Gabriel	J. Pinola	M. Musacchio	Lucas Leiva	J. Rodriguez
D. Ospina	W. Yarbrough	Djavan	F. Coloccini	Rafael	Neymar
Guilherme	M. Andujar	T. Rincon	Marquinhos	C. Ansaldi	A. Di Maria
G. Rulli	M. Goicoechea	C. Gruezo	Juan Jesus	M. Icardi	S. Agüero
Neto	M. Caranta	J. Silva	H. Moreno	Fernando	A. Vidal
C. Vargas	Da Costa	A. Ramalho	C. Zambrano	Mario F	A. Sanchez

Tabla 21: Segunda clasificación 2015, Fuente: Elaboración propia.

11.3. Análisis de Regresión Logística

Al realizar el análisis clúster de los datos en cada año, en todos se crearon tres grupos en los que se conformaron atacantes-mediocampistas, otro conformado por defensas y un último grupo conformado por los porteros; después se realiza una segunda clasificación en cada uno de los tres grupos mencionados anteriormente, dando como resultado que en cada grupo de generaron dos subgrupos donde en uno se encuentran los jugadores con mejores habilidades y otro con los de menos habilidades.

Los últimos subgrupos generados por el segundo análisis cluster se van a categorizar asignándole un nombre a cada subgrupo sin importar en la posición en la que se desempeñan, el subgrupo que contiene los jugadores con buenas habilidades van a ser llamados “Mejores habilidades” y el segundo subgrupo donde los jugadores tienen una menor habilidad se van a llamar “Habilidad normal”, todo eso con el fin de poder realizar un modelo de regresión logística para poder clasificar nuevos jugadores y poder identificar los de “Mejores habilidades”.

Para crear esta nueva clasificación de jugador con mejores habilidades o habilidades normales que va a ser la variable respuesta del modelo se va a tomar en cuenta la clasificación de los años anteriores ya que en las clasificaciones por año ocurrieron casos de jugadores que dentro de los siete años se clasificaron tanto en la categoría Mejores habilidades como en la categoría habilidades normales, esto tiene mucho sentido ya que el rendimiento de los futbolistas por algún suceso puede cambiar, ya sea que un jugador al inicio del periodo no tenga buenas habilidades va a ser clasificado como jugador con habilidades normales pero que con el paso del tiempo sus habilidades aumentaron y después de cierto número de años el jugador ahora es clasificado como jugador con mejores habilidades, o puede ocurrir que un jugador bueno presente una baja en sus habilidades ya sea por una lesión, inactividad o edad, ahora aparece clasificado como habilidad normal y adicional se puede iniciar otro caso en que en los siete años siempre fue clasificado como mejores habilidades a excepción del tercer año que se clasificó como habilidades normales; entonces para poder tomar en cuenta lo mencionado anteriormente a la nueva variable de respuesta de mejores habilidades o habilidad normal se le agrego una nueva condición, que es si el jugador durante los siete años se clasificó como mejores habilidades mayor o igual a cuatro veces se considera como un jugador con mejores habilidades, de lo contrario se considera como un jugador de habilidades normales, para argumentar lo anterior un jugador con buenas habilidades en el fútbol debe tener una buena continuidad durante varios años, como los jugadores Lionel Messi y Cristiano Ronaldo que por más de 10 años han mantenido estable y muy altas sus habilidades, esto también puede significar que el si el jugador mantiene habilidades altas y continuas es porque se lesiona muy poco, lo anterior es una de las cualidades más importantes de un buen jugador.

Al ya tener creada la nueva variable de clasificación se procede a crear el modelo logístico, las variables de entrada son las mismas que se utilizaron para el análisis por componente principales que son las siguientes:

Variable	Significado	Rango
height_cm	Estatura del jugador.	
weight_kg	Peso del jugador.	
overall	Calificación general del jugador.	VARIABLE DE CALIFICACION DE 0 A 100
potential	Calificación esperada del jugador.	
value_eur	Valor del jugador en Euros.	
wage_eur	Sueldo del jugador en Euros.	
attacking_crossing	Capacidad de cruzar el balón de un lado de la cancha a otro.	VARIABLE DE CALIFICACION DE 0 A 100
attacking_finishing	Capacidad de definición de balón.	
attacking_heading_ac	puntería de disparo.	
attacking_short_passi	Pases en distancias reducidas.	
attacking_volleys	Capacidad para realizar voleas.	
skill_dribbling	Capacidad para evadir jugadores rivales.	
skill_curve	Al disparar el balón que tanta toma curva.	
skill_fk_accuracy	Precisión en el pase.	
skill_long_passing	Capacidad de realizar pases a distancias alejadas.	
skill_ball_control	Capacidad que evalúa los desplazamientos con el balón.	
movement_acceleratio	Es el incremento de velocidad de un jugador.	
movement_sprint_spe	Mide la velocidad del jugador.	
movement_agility	Mide que tan rápido es el jugador mientras se gira.	
movement_reactions	Mide la velocidad de reacción.	
movement_balance	Mide la capacidad de equilibrio del jugador.	
power_shot_power	Mide la fuerza del disparo del jugador.	
power_jumping	Mide la capacidad de salto del jugador.	
power_stamina	Mide la resistencia del jugador al chocar.	
power_strength	Mide la fuerza del jugador.	
power_long_shots	Mide la capacidad de disparo desde distancias alejadas.	
mentality_aggression	Mide la agresividad de un jugador.	
mentality_interception	Mide la capacidad de cortar jugadas.	
mentality_positioning	Mide la capacidad de que el jugador se encuentre en su posición correcta.	
mentality_vision	Mide la capacidad del jugador en crear jugadas.	
mentality_penalties	Mide la capacidad que tiene el jugador para cobrar penaltis.	
mentality_composure	Mide el estado del jugador.	
defending_standing_t	Mide la efectividad de quitar el balón sin barrerse.	
defending_sliding_tac	Mide la efectividad de quitar el balón barriéndose.	
goalkeeping_diving	Mide la fuerza del jugador en las manos como arquero.	
goalkeeping_handling	Mide la capacidad de agarre de balón como arquero.	
goalkeeping_kicking	Mide la forma de patear el balón, pero como arquero.	
goalkeeping_positioni	Mide el posicionamiento del jugador como arquero.	
goalkeeping_reflexes	Mide los reflejos del jugador como arquero.	

Tabla 22: Variables a utilizar en la regresión, Fuente: Elaboración propia.

Resumen por variable:

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
height_cm	157,00	175,00	180,00	179,80	185,00	196,00
weight_kg	56,00	71,00	76,00	76,11	80,00	101,00
overall	60,00	70,00	73,00	73,75	77,00	93,00
potential	60,00	71,00	74,00	74,51	78,00	93,00
value_eur	0,00	1100000,00	3200000,00	6181022,00	7125000,00	9000000,00
wage_eur	0,00	5000,00	13000,00	25309,00	26000,00	560000,00
attacking_crossing	11,00	48,75	64,00	57,45	71,00	90,00
attacking_finishing	5,00	38,00	56,00	52,22	68,00	95,00
attacking_heading_accuracy	10,00	54,00	64,00	59,40	72,00	88,00
attacking_short_passing	14,00	64,00	70,00	66,63	75,25	91,00
attacking_volleys	7,00	36,00	55,00	50,81	67,00	90,00
skill_dribbling	7,00	58,00	69,00	62,78	75,00	96,00
skill_curve	8,00	43,00	64,00	56,65	72,00	93,00
skill_fk_accuracy	10,00	36,00	55,00	51,44	68,00	94,00
skill_long_passing	12,00	58,75	65,00	62,17	71,00	91,00
skill_ball_control	11,00	64,00	71,00	66,32	76,00	96,00
movement_acceleration	29,00	56,00	68,00	65,73	77,00	95,00
movement_sprint_speed	28,00	55,00	68,00	65,55	77,00	94,00
movement_agility	28,00	60,75	70,00	67,76	77,25	96,00
movement_reactions	54,00	67,00	70,00	70,90	75,00	94,00
movement_balance	23,00	56,00	67,00	65,32	76,00	95,00
power_shot_power	27,00	60,75	71,00	68,08	76,00	90,00
power_jumping	30,00	65,00	72,00	70,90	79,00	93,00
power_stamina	17,00	62,75	71,00	67,91	78,00	94,00
power_strength	30,00	64,00	72,00	70,62	78,00	94,00
power_long_shots	9,00	46,00	64,00	56,82	71,00	94,00
mentality_aggression	11,00	58,00	70,00	66,09	78,00	95,00
mentality_interceptions	10,00	36,00	64,00	55,02	71,00	89,00
mentality_positioning	7,00	47,75	64,50	57,49	72,00	94,00
mentality_vision	11,00	55,00	65,00	62,42	72,00	95,00
mentality_penalties	12,00	45,00	59,00	55,70	68,00	92,00
mentality_composure	23,00	64,00	70,00	69,20	75,00	96,00
defending_standing_tackle	7,00	35,00	64,00	54,41	72,25	88,00
defending_sliding_tackle	8,00	29,00	61,00	51,70	71,00	87,00
goalkeeping_diving	2,00	8,00	11,00	17,55	14,00	90,00
goalkeeping_handling	2,00	9,00	11,00	17,37	14,00	82,00
goalkeeping_kicking	2,00	9,00	12,00	17,41	15,00	93,00
goalkeeping_positioning	2,00	8,00	11,00	17,55	15,00	86,00
goalkeeping_reflexes	2,00	8,00	11,00	17,73	14,00	90,00

Tabla 23: Resumen variables, Fuente: Elaboración propia.

La base de datos contiene 592 jugadores latinoamericanos de futbol, después de revisar los resúmenes de las variables se puede ver que la mayoría están entre el rango de 0 – 100 a excepción de las variables

height cm, weight kg, value eur y wage eur, la clasificación de los jugadores de la base de datos fue la siguiente:

Mejores Habilidades	Habilidades Normales
366	226.

Tabla 24: Clasificación jugadores, Fuente: Elaboración propia.

Para la elaboración del modelo de regresión logístico se divide base de datos en dos una que contiene el 80 % del total de la información para poder elaborar el modelo y el restante se va a utilizar para probar las predicciones del modelo, de los 592 jugadores del total, se tomaron 474 jugadores para poder crear el modelo y 118 para probar los resultados del modelo.

Primero se crea el modelo con todas las variables y el resumen es el siguiente:

Null deviance	629.42 on 473 degrees of freedom
Residual deviance	456.61 on 434 degrees of freedom.
AIC	536.61

Tabla 25: Estadísticos modelo, Fuente: Elaboración propia.

Variables	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-23,71965	7,43920	-3,18847	0,00143	**
height_cm	0,02703	0,03558	0,75972	0,44742	
weight_kg	0,04650	0,03325	1,39846	0,16197	
overall	0,03975	0,10652	0,37315	0,70903	
potential	0,04422	0,09088	0,48660	0,62654	
value_eur	0,00000	0,00000	-0,81820	0,41325	
wage_eur	0,00001	0,00001	1,82423	0,06812	*
attacking_crossing	0,03568	0,01847	1,93198	0,05336	*
attacking_finishing	-0,02044	0,01936	-1,05575	0,29108	
attacking_heading_accuracy	0,03296	0,01870	1,76278	0,07794	*
attacking_short_passing	-0,00816	0,03440	-0,23726	0,81245	
attacking_volleys	0,01060	0,01563	0,67826	0,49761	
skill_dribbling	0,00979	0,03114	0,31451	0,75313	
skill_curve	-0,00880	0,01570	-0,56077	0,57496	
skill_fk_accuracy	0,00451	0,01354	0,33278	0,73930	
skill_long_passing	0,04669	0,02638	1,76993	0,07674	*
skill_ball_control	0,04142	0,03936	1,05235	0,29264	
movement_acceleration	-0,02718	0,02565	-1,05984	0,28922	
movement_sprint_speed	-0,01134	0,02303	-0,49247	0,62239	
movement_agility	0,01577	0,02086	0,75611	0,44958	
movement_reactions	0,08435	0,03623	2,32824	0,01990	*
movement_balance	0,00216	0,01934	0,11184	0,91095	
power_shot_power	0,02442	0,02179	1,12057	0,26247	
power_jumping	-0,01215	0,01237	-0,98195	0,32613	
power_stamina	0,00270	0,01508	0,17882	0,85808	
power_strength	-0,03214	0,02079	-1,54602	0,12210	
power_long_shots	-0,02313	0,02152	-1,07518	0,28229	
mentality_aggression	0,00869	0,01341	0,64778	0,51713	
mentality_interceptions	-0,01352	0,01948	-0,69398	0,48770	
mentality_positioning	-0,03551	0,02252	-1,57688	0,11482	
mentality_vision	0,03940	0,02219	1,77567	0,07579	*
mentality_penalties	-0,03365	0,01563	-2,15331	0,03129	*
mentality_composure	0,00634	0,02497	0,25390	0,79957	
defending_standing_tackle	-0,02956	0,02839	-1,04105	0,29785	
defending_sliding_tackle	-0,00152	0,02495	-0,06108	0,95129	
goalkeeping_diving	-0,00622	0,03604	-0,17258	0,86298	
goalkeeping_handling	0,05668	0,03899	1,45385	0,14599	
goalkeeping_kicking	0,00221	0,03442	0,06409	0,94890	
goalkeeping_positioning	0,01878	0,03479	0,53974	0,58938	
goalkeeping_reflexes	-0,05868	0,03763	-1,55947	0,11889	

Tabla 26: Resumen modelo con todas las variables, Fuente: Elaboración propia

Al revisar el resumen del modelo se puede observar que de las 39 variables juntas en el modelo solo dos variables son significativas, que se puede interpretar que el modelo completo no funciona adecuadamente y se deben buscar nuevas combinaciones entre variables. Para seguir comprobando que el modelo completo no es viable se trata de realizar una clasificación con los datos de prueba, y se arroja lo siguiente:

Precisión	0,7288136
-----------	-----------

Tabla 27: Precisión modelo, Fuente: Elaboración propia

La precisión de la predicción de clasificación es de aproximadamente el 73 %, La tasa de error de clasificación errónea es del 27 %.

	Mejores Habilidades	Habilidades Normales
Mejores Habilidades	25	11
Habilidades Normales	21	61

Tabla 28: Clasificación jugadores, Fuente: Elaboración propia.

Al ser 39 variables y para poder encontrar el modelo que más se ajuste se procede a utilizar la función Step en Rstudio para seleccionar el modelo basado en el criterio AIC, y también con el algoritmo backward, arrojando como mejor modelo el siguiente:

Step	Df	Deviance	Resid, Df	Resid, Dev	AIC
1	NA	NA	434	456,614051	536,614051
2 <u>defending_sliding_tackle</u>	1	0,00373171	435	456,617783	534,617783
3 <u>goalkeeping_kicking</u>	1	0,00410099	436	456,621884	532,621884
4 <u>movement_balance</u>	1	0,01115911	437	456,633043	530,633043
5 <u>goalkeeping_diving</u>	1	0,02501106	438	456,658054	528,658054
6 <u>power_stamina</u>	1	0,03512609	439	456,69318	526,69318
7 <u>attacking_short_passing</u>	1	0,0482357	440	456,741416	524,741416
8 <u>mentality_composure</u>	1	0,0557041	441	456,79712	522,79712
9 <u>skill_dribbling</u>	1	0,10397613	442	456,901096	520,901096
10 <u>skill_fk_accuracy</u>	1	0,08253235	443	456,983628	518,983628
11 <u>overall</u>	1	0,15394581	444	457,137574	517,137574
12 <u>skill_curve</u>	1	0,24828669	445	457,385861	515,385861
13 <u>movement_sprint_speed</u>	1	0,21977488	446	457,605636	513,605636
14 <u>goalkeeping_positioning</u>	1	0,39610509	447	458,001741	512,001741
15 <u>mentality_interceptions</u>	1	0,47255091	448	458,474292	510,474292
16 <u>attacking_volleys</u>	1	0,49988035	449	458,974172	508,974172
17 <u>height_cm</u>	1	0,51898478	450	459,493157	507,493157
18 <u>movement_agility</u>	1	0,52167776	451	460,014835	506,014835
19 <u>mentality_aggression</u>	1	0,59638467	452	460,611219	504,611219
20 <u>value_eur</u>	1	0,70169524	453	461,312914	503,312914
21 <u>power_jumping</u>	1	0,95570897	454	462,268623	502,268623
22 <u>attacking_finishing</u>	1	1,01749829	455	463,286122	501,286122
23 <u>potential</u>	1	0,89676351	456	464,182885	500,182885

Tabla 29: Selección de variables para el mejor modelo, Fuente: Elaboración propia.

El modelo anterior con las combinaciones que realizó la función Step, es el de menor AIC y con la mayoría

de variables significativas:

Null deviance	629.42 on 473 degrees of freedom
Residual deviance	464.18 on 456 degrees of freedom.
AIC	500.18

Tabla 30: Estadísticos modelo, Fuente: Elaboración propia.

Variables	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-15,638013	2,976151	-5,254442	0,000000
weight_kg	0,058916	0,028929	2,036539	0,041696
wage_eur	0,000013	0,000006	2,216169	0,026680
attacking_crossing	0,033974	0,016179	2,099952	0,035733
attacking_heading_accuracy	0,028917	0,015692	1,842793	0,065359
skill_long_passing	0,050553	0,022378	2,259003	0,023883
skill_ball_control	0,053091	0,030573	1,736551	0,082466
movement_acceleration	-0,030597	0,012410	-2,465426	0,013685
movement_reactions	0,101266	0,027747	3,649562	0,000263
power_shot_power	0,030725	0,020139	1,525645	0,127098
power_strength	-0,036626	0,017999	-2,034914	0,041860
power_long_shots	-0,033348	0,019164	-1,740080	0,081845
mentality_positioning	-0,037972	0,019380	-1,959363	0,050070
mentality_vision	0,042959	0,020190	2,127752	0,033358
mentality_penalties	-0,032257	0,013468	-2,395070	0,016617
defending_standing_tackle	-0,034736	0,010153	-3,421433	0,000623
goalkeeping_handling	0,062914	0,032472	1,937488	0,052686
goalkeeping_reflexes	-0,045127	0,029534	-1,527981	0,126517

Tabla 31: Mejor modelo AIC, Fuente: Elaboración propia.

Las variables del modelo se pueden interpretar de la siguiente forma:

- Entre mayor peso (height cm) tenga el jugador de fútbol, va a tener una mayor probabilidad de ser un jugador de mejor habilidad, se debe aclarar que el peso se relaciona con la fuerza del jugador ya que si tiene bastante peso el jugador va a tener muchas complicaciones para jugar.
- Entre mayor sueldo (wage eur) tenga el jugador de fútbol, va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mayor habilidad de cambios de balón con el pie (attacking crossing) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mayor puntería de disparo (attacking heading accuracy) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.

- Entre mayor habilidad en pases largos (skill long passing) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mayor control de balón (skill ball control) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menor aceleración tenga el jugador (movement acceleration) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mayor velocidad de reacción del jugador (movement reactions) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mayor fuerza de disparo del jugador (power shot power) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menor fuerza del jugador (power strength) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menor fuerza de disparo largo del jugador (power long shots) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menor tenga la habilidad de posición del jugador (mentality positioning) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mejor capacidad de crear jugadas el jugador (mentality vision) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menor capacidad de cobrar penalties del jugador (mentality penalties) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menores habilidades defensivas del jugador (defending standing tackle) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre mejor agarre del balón del jugador (goalkeeping handling) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.
- Entre menores reflejos del jugador tenga (goalkeeping reflexes) va a tener una mayor probabilidad de ser un jugador de mejor habilidad.

Se calculan los intervalos de confianza del modelo con mejor AIC:

	height	2.5 %	97.5 %
(Intercept)		-21.62	-9.92
weight_kg		0.00	0.12
wage_eur		0.00	0.00
attacking_crossing		0.00	0.07
attacking_heading_accuracy		-0.00	0.06
skill_long_passing		0.01	0.10
skill_ball_control		-0.01	0.11
movement_acceleration		-0.06	-0.01
movement_reactions		0.05	0.16
power_shot_power		-0.01	0.07
power_strength		-0.07	-0.00
power_long_shots		-0.07	0.00
mentality_positioning		-0.08	-0.00
mentality_vision		0.00	0.08
mentality_penalties		-0.06	-0.01
defending_standing_tackle		-0.05	-0.02
goalkeeping_handling		-0.00	0.13
goalkeeping_reflexes		-0.10	0.01

Tabla 32: Intervalos de confianza modelo, Fuente: Elaboración propia.

Se realizar con el modelo de mejor AIC una clasificación con los datos de prueba, y se arroja lo siguiente:

Precisión	0,7372881
-----------	-----------

Tabla 33: Precisión modelo, Fuente: Elaboración propia

La precisión de la predicción de clasificación es de aproximadamente el 74 %, La tasa de error de clasificación errónea es del 26 %.

	Mejores Habilidades	Habilidades Normales
Mejores Habilidades	26	11
Habilidades Normales	20	61

Tabla 34: Clasificación jugadores, Fuente: Elaboración propia.

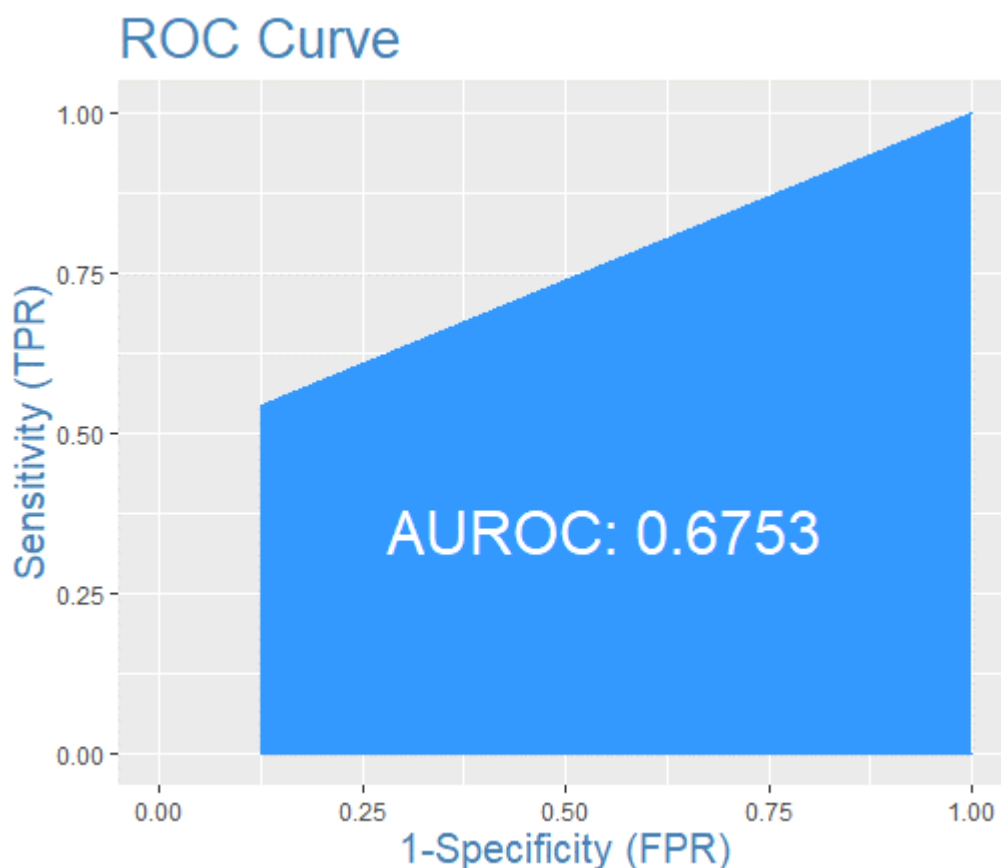


Figura 22: Curva ROC modelo, Fuente: Elaboración propia.

El modelo de regresión logística con menor AIC tiene un gran porcentaje de error, lo cual deja como conclusión presentar un nuevo modelo distinto al logístico para poder realizar una clasificación más acertada.

Por último, se prueba el modelo con los datos del 2021 dando las siguientes clasificaciones:

Short name	age	Nationality	Club Name	League Name	Pred	Clasificación
R. Borre	24	Colombia	River Plate	Argentina Primera División	1	Mejores Habilidades
A Rossi	24	Argentina	Boca Juniors	Argentina Primera División	1	Mejores Habilidades
J. Carrascal	22	Colombia	River Plate	Argentina Primera División	1	Mejores Habilidades
F Troyansky	23	Argentina	Unión	Argentina Primera División	1	Mejores Habilidades
J Rodriguez	24	Argentina	Banfield	Argentina Primera División	1	Mejores Habilidades
M. Ojeda	21	Argentina	Godoy Cruz	Argentina Primera División	1	Mejores Habilidades

Tabla 35: Estimacion jugadores jovenes de la liga argentina, Fuente: Elaboración propia.

En la liga argentina se encuentran 6 jugadores clasificados como buenos menores de 25 años, por críticas de personas relacionadas con el futbol afirman que estos jugadores son muy buenos talentos a futuros, lo cual hace pensar que el modelo realizo una predicción bien.

12. Conclusiones e investigaciones futuras

- Por los resultados del análisis de componentes principales se puede evidenciar que no todas las variables dentro del estudio son importantes, como el caso del sueldo y valor del jugador ya que son variables que tienen una correlación muy baja con las variables físicas y mentales, esto tiene mucho sentido ya que se tienen casos donde hay jugadores que no tienen buenas condiciones o sus habilidades a través de los años han disminuido pero tienen un valor o sueldo muy alto en el mercado ya que por temas de marketing generan mayores ingresos al equipo de futbol.
- El análisis se debe mirar por las habilidades que tiene el jugador mas no por la posición en la que juegan ya que por los clúster realizados se pueden observar jugadores que ocupan posiciones defensivas se clasifican con los que ocupan posiciones de ataque, esto se debe a un cambio que se ha venido presentando en el futbol en los últimos años, ya que la mayoría de jugadores de elite deben tener al tope todas las habilidades físicas y mentales, volviendo al jugador casi perfecto que no tenga debilidades defensivas o de ataque, la única excepción a esto son los jugadores que tienen la posición de arquero ya que son únicos en su posición y por reglamento ellos pueden utilizar las manos, entonces su perfeccionamiento en habilidades van enfocadas a los reflejos.
- Se puede cumplir con una reducción de variables ya que se tienen muchas similares como todas las variables de arquero están muy relacionadas entre si como lo confirma el análisis de componentes principales y el modelo logístico.
- Con respecto a los jugadores latinoamericanos se puede decir que se destacan por tener muy buen control de balón, buenas habilidades de regate, velocidad, cambio de ritmo, resistencia y fuerza, pero no tienen una muy buena preparación física lo cual hace que sean mas propensos a lesiones y no tener una buena continuidad a lo largo de su carrera profesional, esto además de ser resultado del análisis de componentes principales se puede explicar por el contexto geográfico y social, ya que los países latinoamericanos se caracterizan por ser muy acogedores y fiesteros, lo cual puede traer como consecuencia que se tenga una disciplina distinta a Europa, esto afecta bastante a los futbolistas ya que al llevar un ritmo de vida sin disciplina o combinar su carrera con acciones distintas a la profesión que ejercen genera un desgaste físico que es perjudicial en su rendimiento.
- El modelo propuesto cumple con el objetivo y al realizar pruebas con jugadores jóvenes que tienen una buena proyección visual a futuro acierta en la mayoría de casos, pero para poder mejorar la clasificación se puede agregar variables mas específicas, como la cantidad de goles por año, una variable que contenga el promedio de calificación por año, la cantidad de lesiones que tiene durante el año; adicional se pueden probar otros métodos de clasificación para los clúster que pueda detectar jugadores que no tienen habilidades superiores pero son muy reconocidos y juegan en los equipos más importantes del mundo.
- Se puede continuar el estudio con modelos entrenados con herramientas de minería de datos como lo son un modelo Random forest o redes neuronales que realizan mas pruebas en los datos entrenados para obtener mejores predicciones.
- Se recomienda a los equipos de futbol colombiano utilizar estos métodos de predicción de jugadores, esto puede traer muchos ahorros económicos y una mejor elección de jugadores que no solo van a contribuir a mejorar el nivel deportivo del equipo sino del país.

Lagarrañaga, P. (2012). Redes Neuronales. ehu. <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf>.

Idoate, G. (2020, 22 noviembre). Fútbol 11: Qué es, Duración de los Partidos, Posiciones de Juego e Info. Mis entrenamientos de fútbol. [https://www.misentrenamientosdefutbol.com/diccionario/futbol-11: %7E:text=El %20f %C3 %BA t%20bol %2011 %2C %20tambi %C3 %A9 n %20conocido,cuales %20se %20enfrentan %20entre](https://www.misentrenamientosdefutbol.com/diccionario/futbol-11-%7E:text=El%20f%C3%BAt%20bol%2011%2C%20tambi%C3%A9n%20conocido,cuales%20se%20enfrentan%20entre)

colaboradores de Wikipedia. (2021, 9 febrero). Fútbol. Wikipedia, la enciclopedia libre. [https://es.wikipedia.org/wiki/F %C3 %B3 t %20bol](https://es.wikipedia.org/wiki/F%C3%B3t%20bol)

Garcia, D. (2021). Una aplicación de redes neuronales artificiales para el pronóstico de los rendimientos de la serie de Ecopetrol. Repositorio USTA. <https://repository.usta.edu.co/bitstream/handle/11634/31765/2021duvangarcia>.

Gomez, P. (2009). Redes Neuronales Artificiales Conceptos básicos. Inaoep. [https://ccc.inaoep.mx/ pgomez/cursos/redes %](https://ccc.inaoep.mx/pgomez/cursos/redes%20neuronales)

Rodríguez, D. (2018, 1 julio). La regresión logística. Analytics Lane. <https://www.analyticslane.com/2018/07/23/la-regresion-logistica/>.

colaboradores de Wikipedia. (2021a, enero 18). Regresión logística. Wikipedia, la enciclopedia libre. [https://es.wikipedia.org/wiki/Regresi %C3 %B3 n log %C3 %A D stica](https://es.wikipedia.org/wiki/Regresi%C3%B3n_log%C3%ADstica).