

ISSN 1870-4069

Clasificación de jugadores de futbol soccer basada en sus habilidades deportivas, físicas y mentales

Enrique Antonio Pedroza Santiago, Maricela Quintana López,
Héctor Rafael Orozco Aguirre, Víctor Manuel Landassuri Moreno

Centro Universitario UAEM Valle de México,
México

enriquepedroza2012@gmail.com, {mquintanal, hrorozcoa, vmlandassurim}@uaemex.mx

Resumen. Regularmente, para determinar la posición adecuada en la que un jugador de futbol soccer debe jugar, el director técnico realiza diversos movimientos a fin de conocer cómo se desarrolla un jugador en cada posición en la cancha, esto con el objetivo de conocer las habilidades que cada jugador debe desarrollar. Este artículo presenta la propuesta de un clasificador de jugadores de futbol soccer, determinando su posición con base en las habilidades deportivas, físicas y mentales de éstos. Para ello, se utilizará un comparativo de dos algoritmos, uno de generación de árboles de decisión (J48) y otro de generación de reglas (PART), esto con el objetivo de analizar las diferentes habilidades que conllevan a una clasificación, así como determinar cuál de éstos algoritmos tiene una mejor precisión. De los experimentos, se puede concluir que los mejores resultados con datos numéricos lo tiene J48, mientras que PART trabaja mejor con datos nominales. De igual forma, es posible concluir que hay habilidades muy bien definidas en algunas posiciones, como es el caso del portero, mientras que en otras como la del mediocampista no lo están del todo.

Palabras clave: Habilidad, clasificación, minería de datos, jugador, futbol.

Classification of Soccer Players Based on Their Sport, Physical and Mental Skills

Abstract. Regularly, to determine the appropriate position in which a soccer player should play, the technical director performs various movements in order to know how a player develops in each position on the court, this with the objective of knowing the skills that each player must develop. This article presents the proposal of a classifier of soccer players, determining their position based on their sport, physical and mental skills. To do this, a comparison of two algorithms will be used, one for the generation of decision trees (J48) and other for the generation of rules (PART), with the aim of analyzing the different skills that lead to a classification, as well as how to determine which of these algorithms has a better precision. From the experiments, it can be concluded that the best results with numerical data is J48, while PART works better with nominal data.

Similarly, it is possible to conclude that there are very well defined skills in some positions, as in the case of the goalkeeper, while others such as the midfielder are not at all.

Keywords: Skill, classification, data mining, player, soccer.

1. Introducción

El fútbol soccer es un deporte táctico donde participan diversos actores tales como los jugadores, el director técnico y el árbitro, este último es quien se encarga de dirigir los encuentros. Las habilidades personales de los jugadores se consideran fundamentales para un buen rendimiento deportivo y junto con las estrategias del equipo, influyen o determinan el resultado del encuentro a favor o en contra.

Cuando se realizan entrenamientos a nivel profesional o semiprofesional, los jugadores no sólo adquieren y desarrollan como factores sus habilidades físicas y deportivas, sino que también fortalecen sus habilidades mentales, todos estos factores influyen para lograr un buen desempeño y resultado a favor en un encuentro.

Es necesario encontrar un equilibrio en las habilidades de los jugadores, esto con el objetivo de tener un mejor rendimiento tanto en el juego como en el aspecto personal.

De acuerdo con Jover [1], cuando un jugador fortalece sus habilidades mentales, puede mejorar sus capacidades de convivencia, disfruta del deporte que practica y encuentra un nivel adecuado como deportista.

El fútbol es un deporte de conjunto donde cada equipo se conforma de un total de once jugadores en la cancha, cada jugador tiene una posición y función definida, la cual le es asignada por el director técnico, dependiendo de las habilidades mostradas como fortalezas en los entrenamientos. Estas posiciones se dividen principalmente en cuatro: portero, defensa, medio y delantero, con las cuales se trabajaron; aunque dependiendo de qué tan alejado este un jugador de la portería o la dirección donde se encuentre ubicado, estas pueden subdividirse en otras que no son consideradas en este trabajo. Cada posición cuenta con un objetivo el cual se describe a continuación:

- Portero: su función es evitar que el equipo contrario anote. Es el único que puede tomar el balón con las manos. Sólo se permite un portero por equipo.
- Defensa: se encargan de evitar que los jugadores del equipo contrario lleguen a la portería. Una alineación clásica, incluye 4 defensas, aunque pueden variar dependiendo de la formación establecida por el director técnico.
- Medio: se encuentran en la mitad de la cancha. Apoyan tanto en la defensa como en el ataque. Su función principal es el de distribuir los balones.
- Delantero: debido a su posición, son quienes anotan principalmente los goles.

Regularmente, en los entrenamientos a los jugadores se le asignan posiciones a modo de probar cómo se desenvuelven en cada una de ellas, de manera que aquella donde cada jugador se sienta más cómodo y dé mejores resultados sea en la que juegue de forma constante o permanente en los encuentros. Las habilidades de un jugador determinan la posición en la que obtendrá un mejor desempeño.

Una manera de apoyar en la toma de decisiones respecto a cuál es la posición más adecuada para un jugador, basada en sus habilidades, es el tener un mecanismo

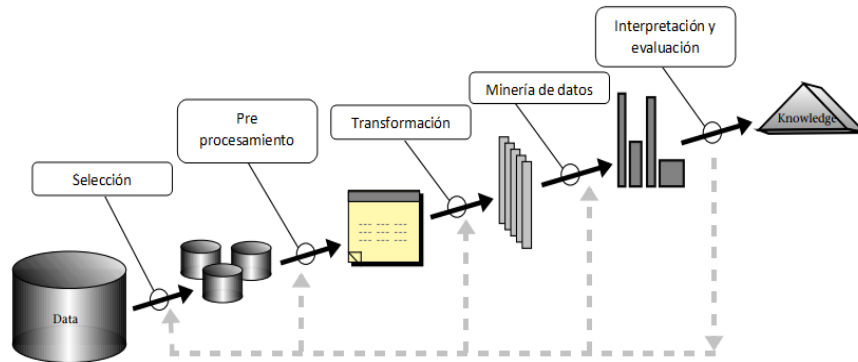


Fig. 1. Etapas de la extracción del conocimiento.

confiable de clasificación. Existen clasificadores de diferentes tipos, algunos de caja negra como las redes neuronales artificiales, que realizan la clasificación pero el conocimiento adquirido no es fácil de examinar [2] y los de caja blanca, como los generadores de árboles de decisión o reglas de clasificación, en los que se puede analizar la estructura que se forma al adquirir el conocimiento. Los algoritmos a utilizar son el C4.5, que genera árboles de decisión y permiten determinar qué atributos se seleccionan para obtener una clasificación y, el algoritmo PART de generación de reglas; esto con el objetivo de analizar las diferentes habilidades que conllevan a una clasificación, así como determinar cuál de éstos algoritmos tiene una mejor precisión. El software que se utilizará es Weka de la universidad de Waikato [3].

El resto de este trabajo está organizado como sigue: en la sección 2, se describe la metodología utilizada; mientras que en la sección 3 se incluyen los algoritmos empleados; posteriormente, en la sección 4, se describen los experimentos realizados y sus resultados; finalmente, en la sección 5, se presentan las conclusiones y el trabajo futuro.

2. Metodología

Para la realización de este trabajo, se utilizó la metodología KDD (Knowledge Discovery in Databases), conocida como el proceso de extracción del conocimiento, propuesta por Fayyad [4], ver figura 1. El proceso consiste de cinco etapas secuenciales e iterativas.

Todas las etapas de la extracción del conocimiento son iterativas, debido a que la salida de una de ellas puede retornar a una etapa anterior, esto con la finalidad de depurar los datos correctamente. A continuación se describe en que consiste cada una.

Preparación de los datos. Esta fase incluye las etapas de selección, procesamiento y transformación. Se determina qué datos se necesitan y se selecciona la información que sea relevante y útil, descartando la mayor cantidad de datos erróneos o que no tengan una aportación, este proceso es conocido como selección y limpieza. Posteriormente se

realiza un agrupamiento de todas las características con la finalidad de conocer si existen anomalías o datos redundantes. Finalmente, la transformación de datos se refiere a todo aquel proceso en el cual se modifiquen éstos.

Minería de datos. Una vez seleccionados los datos, se determina aquellos patrones que se desean obtener, el predecir algún comportamiento o el obtener una clase. Para ello se elige el algoritmo que mejor se adapte a los datos.

Interpretación y evaluación. Finalmente en esta etapa se obtienen los resultados y se determina si se logró una evaluación exitosa. Para ello es necesario una etapa de pruebas donde los datos estén fuera de aquellos con los que se entrenó, de manera que si los resultados tienen un porcentaje alto de acierto se puede determinar que son confiables.

En este trabajo, la preparación de los datos se describe en la sección 3, mientras que las de minería de datos y la de interpretación y evaluación se describen en las secciones 4 y 5 respectivamente.

3. Preparación de los datos

La Federation International de Football Association (FIFA) [5], es la encargada de reunir y gobernar a las diversas asociaciones de futbol existentes en el planeta, es además quien organiza los diversos campeonatos internacionales y, es quien tiene la facultad de crear o modificar las reglas del juego. Este organismo reconoce y almacena las distintas habilidades que un jugador profesional debe poseer siendo un total de 34. Estas habilidades incluyen las deportivas, las físicas y las mentales. A estas se les asignan un valor numérico del 0 al 100 dependiendo de la evaluación de cada jugador.

En la tabla 1, se pueden observar los valores de la media y desviación estándar de las habilidades de un jugador profesional reconocidas por la FIFA.

Tabla 1. Habilidades deportivas, físicas y mentales de los jugadores.

DEPORTIVAS			FÍSICAS	MENTALES
CONTROL DEL BALÓN $\bar{x}=59.70, \sigma=17.25$	RETÉN DEL BALÓN $\bar{x}=17.87, \sigma=19.11$	FUERZA DE TIRO $\bar{x}=58.83, \sigma=18.08$	ACELERACIÓN $\bar{x}=66.30, \sigma=14.54$	AGRESIVIDAD
REGATES $\bar{x}=56.62, \sigma=19.62$	DESPEJE $\bar{x}=18.07, \sigma=18.60$	REMATES $\bar{x}=47.26, \sigma=20.18$	ENERGÍA $\bar{x}=65.54, \sigma=16.40$	REACCIONES
CENTROS $\bar{x}=51.78, \sigma=18.77$	REFLEJOS $\bar{x}=18.20, \sigma=20$	TIROS LARGOS $\bar{x}=50.61, \sigma=20.33$	FUERZA $\bar{x}=65.01, \sigma=12.24$	POSICIÓN DE ATAQUE
PASE CORTO $\bar{x}=59.96, \sigma=14.27$	MARCAJE $\bar{x}=45.76, \sigma=21.78$	CURVA $\bar{x}=50.10, \sigma=18.76$	BALANCE $\bar{x}=66.11, \sigma=12.42$	INTERCEPCIÓN
PASE LARGO $\bar{x}=54.80, \sigma=15.52$	BARRIDA $\bar{x}=46.60, \sigma=21.99$	PRECISIÓN TIRO LIBRE $\bar{x}=44.71, \sigma=18.11$	VELOCIDAD DE SPRINT $\bar{x}=66.20, \sigma=14.79$	VISION
COLOCACIÓN $\bar{x}=18.38, \sigma=19.39$	ENTRADA LIMPIA $\bar{x}=47.81, \sigma=22.33$	PENALES $\bar{x}=53.44, \sigma=16.59$	AGILIDAD $\bar{x}=64.25, \sigma=14.20$	COMPOSTURA
ESTIRADA $\bar{x}=18.15, \sigma=19.69$	CABECEO $\bar{x}=54.07, \sigma=18.74$	VOLEAS $\bar{x}=45.69, \sigma=18.74$	SALTO $\bar{x}=67.63, \sigma=10.84$	

Por otro lado, existen un total de 18 equipos en liga mexicana torneo clausura 2018 (ver tabla 2), cada uno tiene de 20 a 40 jugadores registrados ante la Federación Mexicana de Futbol (FMF) [6], la mayoría de ellos no tienen actividad de forma profesional ya que están en preparación constante para poder debutar. De aquellos que tienen actividad se buscaron sus habilidades en la página oficial de la FIFA.

Tabla 2. Equipos participantes en la liga mexicana torneo clausura 2018.

ATLÁS	MONTERREY	SANTOS
AMÉRICA	MORELIA	TIJUANA
CRUZ AZUL	NECAXA	TOLUCA
GUADALAJARA	PACHUCA	U.A.N.L
LEÓN	PUEBLA	U.N.A.M
LOBOS BUAP	QUERÉTARO	VERACRUZ

Se trabajó con un total de 513 jugadores de primera división en México. Con estos datos se formaron dos conjuntos: el de entrenamiento, y el de prueba. Para el primer conjunto se tomaron los datos de los titulares y suplentes de los equipos del torneo clausura 2018 con un total de 440 jugadores, teniendo 52 porteros, 101 delanteros, 130 defensas y 157 medios, mientras que para la prueba se tomó una muestra de 73 jugadores que militaban en la liga mexicana en el torneo apertura 2017 con la restricción de que no se encuentren en los datos del conjunto de entrenamiento. Para ello se trabajo con 8 porteros, 18 delanteros, 26 defensas y 21 medios; cada uno de ellos con 34 habilidades, las cuales fueron mencionadas en la tabla 1.

Los algoritmos usados admiten datos tanto numéricos como nominales, por ello se hicieron dos análisis, uno utilizando datos con escala de 0-100, mientras que en el segundo se utilizó una escala Likert de 3 rangos: bajo, medio y alto. Esta transformación de datos se realizó con el objetivo de tener una mejor comprensión de los mismos, donde es más fácil asimilar que un jugador es bueno por tener habilidades medias o altas, que en el caso donde se tiene una evaluación numérica.

4. Minería de datos

La minería de datos es el proceso mediante el cual es posible obtener información precisa y valiosa de aquellos datos que se encuentran desorganizados [7]. Existen investigaciones donde se han utilizado algoritmos de minería de datos en el deporte, ejemplo de ellos es [8], donde es posible evaluar el rendimiento de los jugadores de voleibol utilizando los algoritmos C4.5 y EM. Otro ejemplo es el encontrado en [9], donde utilizan el clasificador Naive Bayes para obtener un pronóstico en los encuentros de tenis. Otro enfoque que han utilizado es para el comportamiento de las apuestas deportivas, específicamente en quinielas como lo es el caso de [10].

Existen diversos algoritmos de clasificación, entre ellos se encuentran ID3 y C4.5, los cuales suelen ser precisos y dan una gran confiabilidad [11], este último es el que se utilizó para determinar las posiciones de cada jugador.

Otra opción para clasificar es la generación de reglas, para construirlas se encuentran relaciones entre diversos atributos, como es el caso de PART.

4.1. Algoritmo C4.5 (J48)

C4.5 también conocido como J48 es un algoritmo utilizado para resolver problemas de decisión o de clasificación [12]. Este surge en el año de 1993 por J. Quinlan, como

una mejora del algoritmo ID3, el cual solo admite valores numéricos. C4.5 genera árboles al encontrar un atributo que tenga la mayor ganancia y es utilizado como nodo raíz, posteriormente la división de datos se basa en la recursividad de estos. Una de las características principales de los arboles generados por C4.5 es el radio de ganancia, el cual considera el número de nodos que tiene el árbol y de esta manera el atributo raíz puede dividir los demás conjuntos sin importar la información de la clase. Otro de los aspectos es la poda de árbol que permite que no se expanda cuando los datos se repiten o no son relevantes.

4.2. PART

Es un algoritmo de clasificación basado en reglas, creado por Witten y Frank en el año de 1998 [13]. Este algoritmo podría considerarse como una mezcla de árboles de decisión y reglas de clasificación. PART adopta una estrategia similar a la de J48 al usar la técnica “divide y vencerás”, con lo cual crea las reglas al tomar las ramas que tengan una mayor cobertura y elimina aquellas que no cumplan con las condiciones dadas. Estas se siguen creando de manera recursiva hasta que no queden atributos a considerar. Una de las ventajas que tiene PART sobre otros algoritmos como PRISM [14] es que sus reglas son muy cortas y toma sólo los atributos más relevantes.

5. Experimentos y resultados

El primer experimento con el conjunto de datos numéricos de entrenamiento se realizó aplicando el algoritmo C4.5. Los resultados son presentados en la tabla 3.

Tabla 3. Resultados del algoritmo C4.5 con el conjunto de entrenamiento de datos numéricos.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	64	1	13	23	101	63.4	36.6
PORTERO	4	41	4	3	52	79	21
DEFENSA	8	6	92	24	130	70.8	29.2
MEDIO	11	12	16	118	157	75.2	24.8
Matriz de confusión 315 en la diagonal					440	71.59	28.41

Se obtuvo un 71.59% de instancias clasificadas correctamente, lo que equivale a un total de 315 jugadores de 440. Interpretando la tabla anterior, la primera posición que aparece es la de delantero. Esta aparece con 64 jugadores clasificados correctamente, mientras que existe confusión en un caso con el portero, 13 con la defensa y 23 con la posición medio, esto sucede ya que existen jugadores que comparten habilidades que son muy parecidas en sus valores.

La confusión más grande ocurre en la clase medio, esto se debe a que la función de estos últimos es apoyar tanto en la defensa como en el ataque. La posición de portero es aquella que se encuentra mejor definida y que tiene poca confusión con otras.

El árbol de decisión generado se muestra en las figuras 2 y 3. En ellas, se pueden observar las cualidades del jugador que proporcionan mayor información para separar las clases. Los atributos se visualizan en un óvalo, mientras que las clasificaciones se encuentran en un rectángulo.

Tabla 4. Resultados del algoritmo C4.5 con el conjunto de prueba con datos numéricos.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	16	0	0	2	18	88.9	11.1
PORTERO	0	8	0	0	8	100	0
DEFENSA	2	0	20	4	26	76.9	23.1
MEDIO	4	1	2	14	21	66.7	33.3
Matriz de confusión 58 en la diagonal					73	79.45	20.55

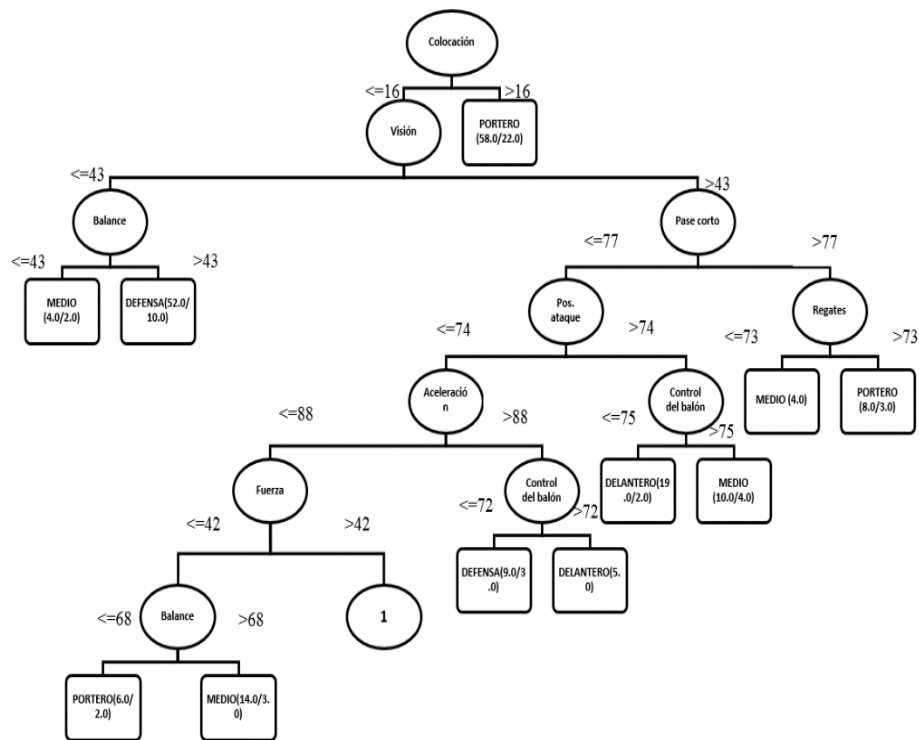


Fig. 2. Primera parte árbol de decisión con datos numéricos.

Analizando el árbol de las figuras 2 y 3, el cual contiene datos numéricos, es posible observar que en la raíz se encuentra la habilidad “colocación”, la cual si es mayor a 16 se determina automáticamente que es un portero, mientras que si es menor se observan otros atributos hasta llegar a una clasificación.

En el proceso de prueba, 58 jugadores fueron clasificados correctamente, lo que corresponde a un 79.45%, mientras que 15 de ellos tuvieron un error, dando un total de 20.55%. Los resultados se pueden observar en la tabla 4.

El segundo experimento realizado fue mediante el algoritmo PART. Dando un 72.73% de acierto y un 27.27% de error. Se obtuvieron un total de 30 reglas de las cuales se observa que las habilidades que más destacan por cada posición son: colocación para el portero, retén del balón para el medio, visión y barrida en el caso de la defensa y aceleración y precisión de tiro en cuestión de los delanteros. Al comparar

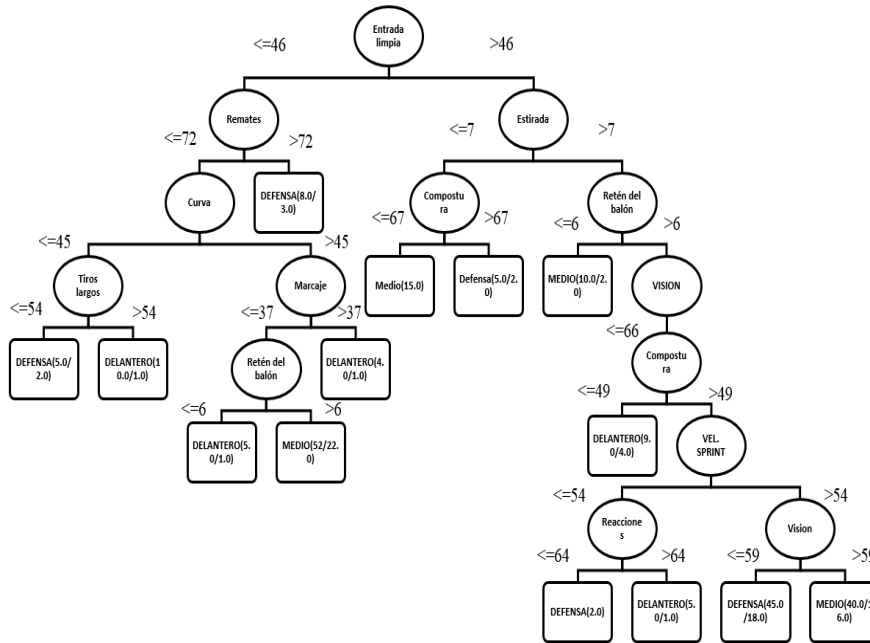


Fig. 3. Segunda parte árbol de decisión con datos numéricos.

Tabla 5. Resultado del algoritmo PART con el conjunto de entrenamiento con datos numéricos.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	58	1	14	28	101	57.4	42.6
PORTERO	4	41	4	3	52	79	21
DEFENSA	4	6	91	29	130	70.0	30.0
MEDIO	7	12	8	130	157	82.8	17.2
Matriz de confusión 320 en la diagonal					440	72.73	27.27

ambos algoritmos se puede determinar que coinciden en las características que debe tener un jugador, teniendo como regla principal la de la colocación, donde el 79% de estos caen en el apartado portero (ver en tabla 5). Del total de las reglas obtenidas, la tabla 6 muestra aquellas con una mayor cobertura y precisión.

Para el proceso de prueba se hizo uso de los mismos 73 jugadores que se usaron con C4.5. Del 100%, el 73.97% de los jugadores tuvieron una clasificación correcta, mientras que en el 26.03% hubo algún error. Los resultados se muestran en la tabla 7.

En la sección 2, se mencionó que los valores numéricos fueron transformados en nominales, para ello se encontraron los valores máximos y mínimos por cada habilidad. Una vez hecho esto, se calcularon los rangos de estos límites.

Posteriormente, se dividió el rango entre 3 valores, y cada resultado recae sobre una clasificación, los cuales son bajo, medio y alto, esto con la finalidad de agrupar valores cercanos y conocer que tan bueno es un jugador, diferente a los valores numéricos donde son específicos para cada uno de estos.

Tabla 6. Conjunto de reglas de clasificación datos numéricos.

Si colocación > 16 y centros > 10 y regates > 17 y remates <= 13.	Entonces = DEFENSA (12.0/4.0)
Si reten del balón <= 16 y visión <= 43 y balance > 43 y regates > 27.	Entonces= DEFENSA (46.0/6.0)
Si reten del balón > 16.	Entonces= PORTERO (46.0/14.0)
Pos. Ataque > 74 y pase largo <= 66.	Entonces: DELANTERO (21.0/3.0)
Si reacciones <= 45 y pase largo <= 52.	Entonces= MEDIO (6.0)
Si intercepción > 30 y visión > 65 y reacciones > 69.	Entonces= MEDIO (35.0/14.0)
Si intercepción > 30 y barrida > 68 y reflejos <= 14 y pase largo <= 72 y estirada <= 10.	Entonces= MEDIO (31.0/14.0)
Si intercepción > 30 y barrida <= 69 y reten > 6 y tiro libre > 59 y retén del balón > 7 y boleas > 45 y visión <= 72.	Entonces= MEDIO (27.0/2.0)
Si entrada > 51 y barrida > 69.	Entonces= DEFENSA (15.0/2.0)
Si aceleración > 81 y despeje > 10.	Entonces= DELANTERO (10.0/1.0)
Si reacciones <= 46 y control > 63.	Entonces= PORTERO (3.0/1.0)
Si colocación > 7 y entrada <= 51 y reten > 7 y centros <= 66 y control > 56 y estirada <= 15.	Entonces= MEDIO (20.0/7.0)
Si colocación > 6 y retén del balón > 6 y estirada <= 9 y compostura <= 63.	Entonces= MEDIO (13.0/2.0)
Si reten del balón > 6 y colocación > 6 y estirada > 9 y entrada > 66 y reflejos <= 13 y Aceleración > 53.	Entonces= DELANTERO (14.0/5.0)
Si colocación > 7 y reten del balón > 6 y aceleración <= 75 y visión > 45 y penales <= 64.	Entonces= MEDIO (21.0/10.0)

Tabla 7. Resultados del algoritmo PART con el conjunto de prueba de datos numéricos.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	10	1	1	6	18	55.6	44.4
PORTERO	0	8	0	0	8	100	0
DEFENSA	1	1	21	3	26	80.8	19.2
MEDIO	3	1	2	15	21	71.4	28.6
Matriz de confusión 54 en la diagonal					73	73.97	26.03

Tabla 8. Resultados del algoritmo C4.5 con el conjunto de entrenamiento de datos nominales.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	70	1	9	21	101	69.3	30.7
PORTERO	4	42	4	2	52	81	19
DEFENSA	14	6	95	15	130	73.1	26.9
MEDIO	9	11	23	114	157	72.6	27.4
Matriz de confusión 321 en la diagonal					440	72.95	27.05

Se creó una base de datos con los valores nominales y se utilizó el algoritmo C4.5, sobre el conjunto de entrenamiento, dando como resultado un 72.95%. Hay un porcentaje parecido de instancias clasificadas correctamente que las obtenidas por los valores numéricos. La matriz de confusión del entrenamiento de datos nominales con el algoritmo C4.5 se puede observar en la tabla 8.

El árbol cambia un poco con relación al creado con los valores numéricos, es mucho más amplio y se evalúan diversos atributos para la clasificación. En este caso toma como nodo raíz la habilidad Estirada.

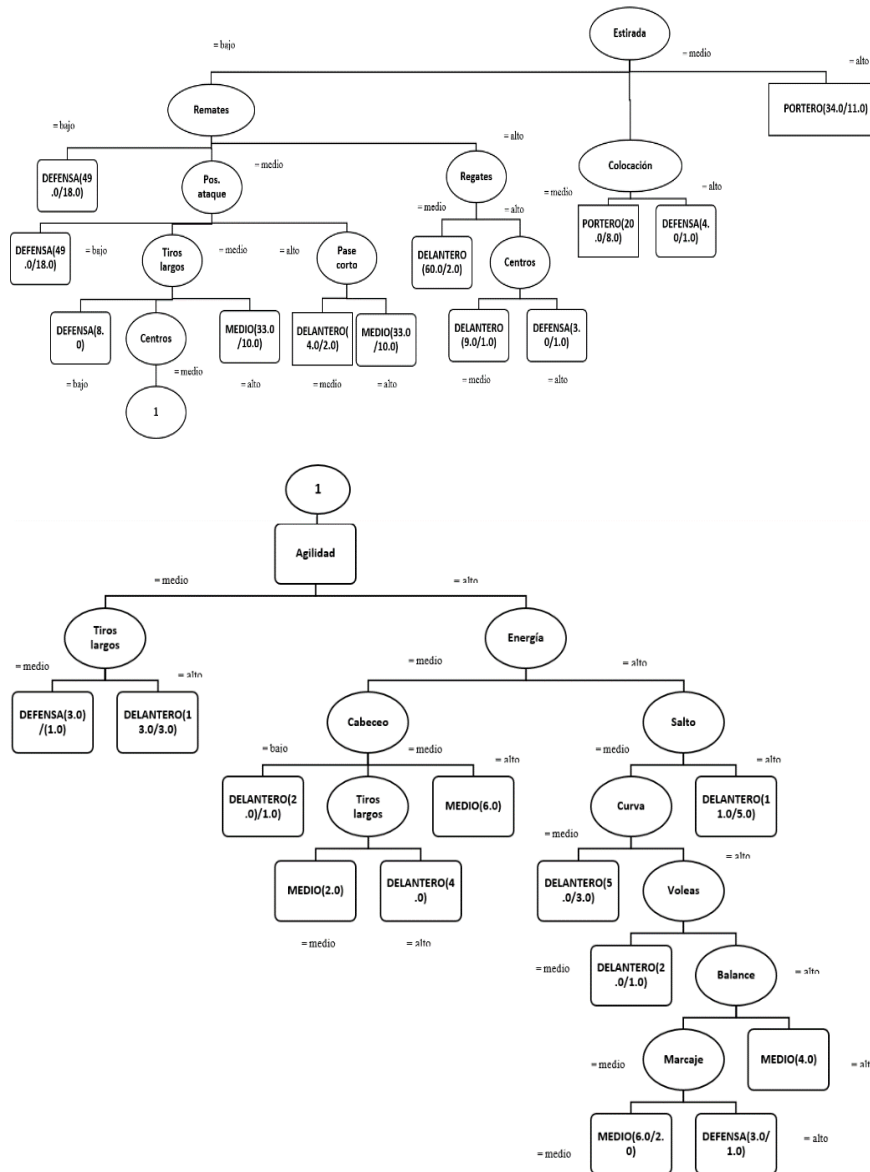


Fig. 4. Árbol de decisión con datos nominales.

Al realizar el proceso de prueba con los datos de los jugadores del torneo pasado se encontraron los siguientes resultados: 50 jugadores que pertenecen al 68.49%, fueron clasificados de forma correcta, los cuales pueden observarse en la matriz de confusión en la tabla 9.

En la cual si su valor es bajo se compara los remates, si es medio, compara la colocación, y si esta es alta se determina automáticamente que es un portero.

Tabla 9. Resultados del algoritmo C4.5 con el conjunto de prueba utilizando datos nominales.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	13	0	2	3	18	72.2	27.8
PORTERO	1	6	1	0	8	75	25
DEFENSA	0	5	19	2	26	73.1	26.9
MEDIO	2	1	6	12	21	57.1	42.9
Matriz de confusión 50 en la diagonal					73	68.49	31.51

Tabla 10. Resultados del algoritmo PART con el conjunto de entrenamiento datos nominales.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	63	1	13	24	101	62.4	37.6
PORTERO	4	41	5	2	52	79	21
DEFENSA	6	6	105	13	130	80.8	19.2
MEDIO	10	11	21	115	157	73.2	26.8
Matriz de confusión 324 en la diagonal					440	73.64	26.36

Tabla 11. Conjunto de reglas con valores nominales utilizando el algoritmo PART.

Si retén del balón = alto y pase corto = bajo y estirada = alto y fuerza = alto.	Entonces= PORTERO (14.0/5.0)
Si colocación = alto y pase corto = bajo y fuerza = medio y salto = medio y reacciones = alto.	Entonces= PORTERO (8.0/2.0)
Si regates = medio y entrada = alto y visión = bajo y vel. sprint = medio y agresividad = alto.	Entonces= DEFENSA (13.0)
Si barrida = alto y pase corto = medio.	Entonces= DEFENSA (27.0/10.0)
Si intercepción = medio y barrida = alto.	Entonces= MEDIO (15.0/4.0)
Si control = medio y visión = medio y compostura = medio.	Entonces= DELANTERO (22.0/10.0)
Si barrida = alto y pase largo = medio.	Entonces= DEFENSA (18.0/6.0)
Si barrida = alto y marcaje = alto y compostura = medio y aceleración = medio.	Entonces= MEDIO (5.0)
Si barrida = alto y compostura = alto y barrida = alto y compostura = alto y balance = alto y penales = medio.	Entonces= MEDIO (14.0/5.0)
Si barrida = alto y compostura = alto y centros = alto.	Entonces= DEFENSA (14.0/5.0)
Si agresividad = medio y barrida = bajo y compostura = alto y marcaje = bajo.	Entonces= DELANTERO (12.0/2.0)
Si agresividad = medio y agilidad = alto y barrida = medio.	Entonces= MEDIO (21.0/8.0)

Tabla 12. Resultados del algoritmo PART con el conjunto de prueba de datos nominales.

CLASIFICACIÓN	DELANTERO	PORTERO	DEFENSA	MEDIO	TOTAL	ACIERTO %	ERROR %
DELANTERO	13	0	2	3	18	72.2	27.8
PORTERO	1	6	0	1	8	75	25
DEFENSA	1	3	21	1	26	80.8	19.2
MEDIO	2	0	7	12	21	57.1	42.9
Matriz de confusión 52 en la diagonal					73	71.23	28.77

El atributo colocación es muy parecido al árbol generado con datos numéricos, ya que los resultados son los mismos, se puede determinar que es defensa o es portero (ver figura 2). Esta es sólo una de las hojas donde se puede analizar las habilidades para determinar la posición de juego. Este árbol se puede observar en la figura 4.

Al igual que con los datos numéricos, se crearon reglas de clasificación utilizando el algoritmo PART, al realizar el entrenamiento mostró un 73.64% de acierto, dando un total de 50 reglas. La matriz de confusión se muestra en la tabla 10 y las reglas que más destacan se encuentran en la tabla 11.

Al realizar la fase de pruebas con los datos de los jugadores del torneo pasado se obtuvo un 71.23% de instancias clasificadas correctamente. Este resultado junto con la matriz de confusión pueden observarse en la tabla 12.

6. Conclusiones y trabajo futuro

El fútbol es un deporte en donde se ven inmersas habilidades tanto deportivas, mentales como físicas, las cuales cada jugador tiene que desarrollar y mejorar día con día. En este trabajo se presentaron dos modelos, uno utiliza árboles de decisión y el otro reglas de clasificación para determinar la posición en la que un jugador de fútbol soccer tendrá posiblemente, un mejor desempeño con base en las habilidades que tiene o demuestra.

Este modelo puede ser de gran ayuda si es utilizada en pruebas o visorias, así como en los entrenamientos en los equipos tanto profesionales como amateur. Se presentaron los modelos con datos numéricos y nominales, que pueden usarse dependiendo de la precisión de la información que se tenga al momento de clasificar al jugador.

En el caso de los datos numéricos se pudo observar que se tiene una mayor certeza con el algoritmo C4.5 teniendo un 71.59% de instancias clasificadas correctamente al realizar el entrenamiento, y un 79.45% en la fase de pruebas, además el árbol de decisión es relativamente corto y tiene mejor definidas las clases, mientras que el algoritmo PART con un 73.64% de certeza en la fase de entrenamiento y un 71.23%, trabaja mejor con datos nominales donde sus reglas son cortas y definen que tan bueno debe ser un jugador en ciertas habilidades para determinar su posición. Ambos algoritmos muestran que un portero sobresale por tener una mejor colocación y reflejos, mientras que la defensa se caracteriza por una mejor compostura y tener mejores tiros largos. La posición medio tiene una mejor visión del campo y retención del balón. Finalmente los delanteros puede observarse que su fortaleza es la aceleración, el cabeceo y el salto.

Se considera que se cumplieron los objetivos planteados, al poder generar un clasificador que permita a un entrenador o director técnico apoyarse para definir la posición de juego de un jugador dándole certeza de las funciones que debe realizar, evitando con ello el tener que probarlo en todas las posiciones tratando de encontrar la mejor para éste.

Se plantea como trabajo futuro mejorar el porcentaje de precisión de la clasificación de jugadores, ya sea tomando en cuenta otros factores tales como la complexión física de estos, el rendimiento que pueda llegar a tener en los entrenamientos, así como el experimentar con otros algoritmos de minería de datos, o en su defecto con otras técnicas de inteligencia artificial.

Referencias

1. Jover, F.: Hábitos de entrenamiento y lesiones deportivas en la selección murciana de baloncesto 2007. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 8(1), pp. 146–160 (2008)

2. Huang, K., Chang, W.: A neural network method for prediction of 2006 world cup football game. In: 2010 International Joint Conference on Neural Networks (IJCNN), IEEE World Congress on Computational Intelligence, 1, pp. 1–8 (2010)
3. WEKA: <http://www.cs.waikato.ac.nz/ml/weka/> (2018)
4. Fayyad, U., Piatetsky-Shapiro, G., Smith, P.: From data mining to Knowledge Discovery and Data Mining. *A.I Magazine*, 17(1), pp. 37–54 (1996)
5. FIFA: <http://es.fifa.com/> (2018)
6. FMF: <http://www.femexfut.org.mx> (2018)
7. Ruiz, R.: Minería de datos como soporte a la toma de decisiones empresariales en una arquitectura SOA, Barranquilla. Ed. Coruniamericana, 1(1), pp. 25–34 (2013)
8. Medina, J., Sandi-Pinheiro, M., Andux, C.: Evaluación del rendimiento de los voleibolistas mediante minería de datos. *Revista Ingeniería industrial, CUJAE*, 26(2), pp. 47–52 (2005)
9. Sanhueza, R.: Utilización de Naive Bayes para predicción de victorias en jugadores de la Asociación de Tenistas Profesionales. Memorias de pregrado Ingeniería civil en computación, Universidad de Talca, Chile (2014)
10. Pérez, F.: Sistema de predicción de apuestas deportivas: una aproximación a la Quiniela. Tesis de grado, Universidad Carlos III, De Madrid, España (2014)
11. Quinlan, J., Kumar, V., Wu, X.: The Top 10 algorithms in data mining. 14(1), pp. 1–37 (2008)
12. Berry, M., Gordon, L.: *Data mining Techniques*. Canada: Wiley Computer Publishing (2004)
13. Ramos, M.: Estudios en finanzas y contabilidad: España y América Latina. Estado del arte y las nuevas metodologías aplicadas, 15, pp. 328–351, Madrid, España (2014)
14. Robles, Y., Sotolongo, A.: Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL. *Journal of Information Systems and Technology Management*, 10(1) (2013)