

EVALUACIÓN DEL RENDIMIENTO DE LOS VOLEIBOLISTAS MEDIANTE MINERÍA DE DATOS

Resumen / Abstract

Para lograr un lugar cimero en el deporte de alta competición, al deportista suele desarrollársele sus facultades motrices, afectivo-sociales y cognitivas. En particular, el aspecto psicológico en este medio resulta de extrema importancia, ya que las acciones motoras dependen, de algún modo, de estas características. Resulta de gran importancia para el entrenador el poder apreciar el comportamiento psicológico de sus atletas en condiciones competitivas para crear medidas y tareas que permitan regular su evolución. En el medio deportivo, más específicamente en el voleibol, la dinámica es muy grande. Por ello, el proceso de análisis de estas informaciones resulta un gran reto científico-técnico. En este trabajo se propone una metodología para realizar el diagnóstico del rendimiento de los voleibolistas en condiciones competitivas, a partir de las variables psicológicas, usando las técnicas de la minería de datos para la realización del análisis. Los resultados obtenidos no solo validaron las hipótesis tenidas, sino que han promovido un estudio más profundo en esta área con los especialistas de este deporte.

To achieve a significant place in the sport of high competition, it is accustomed to develop the motive, affective-social, and cognitive abilities of the sportsman. In particular, the psychological aspect in this environment is of extreme importance, since the motive actions depend, somehow, of these characteristics. It is of great importance for the trainer to be able to appreciate the psychological behaviour of their athletes under competitive conditions to create measures and tasks that allow regulating their evolution. In the sport environment, more specifically in volleyball, the dynamics is very high and the lost of such information is unavoidable. For this reason, the process of analysis of this information is a great scientific-technique challenge. In this paper it is proposed a methodology to carry out the diagnosis of the performance of the volleyball sportsmen under competitive conditions, starting from the psychological variables, using Data Mining techniques to accomplish the analysis. The results obtained didn't only validate the hypotheses, but rather it has promoted a deeper study in this area with the specialists of this sport.

José E. Medina Pagola, Licenciado en Cibernética Matemática, Doctor en Ciencias Técnicas, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV), Ciudad de La Habana, Cuba

e-mail: medinajose@mixmail.com

Marcello Sandi Pinheiro, Licenciado en Análisis de Sistemas, Máster en Informática Aplicada, Universidad Católica de Brasilia, Brasil

e-mail: msandipinheiro@yahoo.com.br

Calixto Andux Despachelles, Licenciado en Educación Física, Doctor en Ciencias Pedagógicas, Instituto Superior de Cultura Física Manuel Fajardo (ISCFMF), Ciudad de La Habana, Cuba

e-mail: andux@inder.co.cu

Palabras clave / Key words

Minería de datos

Data mining, KDD

INTRODUCCIÓN

Múltiples han sido las aplicaciones del *Data Mining* (minería de datos), como son las realizadas para el marketing selectivo, pronóstico financiero, diagnóstico médico, entre otras.

En el área deportiva, una de las aplicaciones paradigmáticas del *Data Mining* (DM) es el *Advanced Scout*. Este sistema fue desarrollado por IBM para la Asociación Nacional de Baloncesto Norteamericano (NBA). Según Tom Sterner, entonces director asistente del Orlando Magic, y director del comité técnico de la NBA:¹

Durante los juegos en la serie de Atlanta Hawks, usamos minería de dato para comprender mejor el problema con los rebotes. Realmente, nos ayudó con nuestro esquema de sustitución.

El talento toma un largo camino para ganar. Si usted tiene más talento que su oponente, ganará. Cuando su talento lo iguala, entonces la arista competitiva pudiera ser la tecnología que use para prepararse para el juego.

También en el área del entrenamiento deportivo, un estudio interesante fue desarrollado por el doctor Lynn Fielitz sobre la minería de datos aplicada al proceso de evaluación física de los cadetes de la Academia Militar de los Estados Unidos.²

Sin embargo, y a pesar de la significación tenida por el sistema Advanced Scout, es de notar que en los últimos años han sido mínimos los trabajos realizados en el área deportiva.

El deporte, especialmente el de alto rendimiento, se ha visto inmerso en un acelerado desarrollo y exigencia que demanda cada vez mayor apoyo técnico y científico. Hoy los deportistas sostienen retos y alcanzan metas que hacía algunos años atrás parecían sueños. Para ello, es necesario que se sometan a rigurosos entrenamientos, los cuales cuentan con el concurso de numerosas disciplinas científicas, si desean llegar a la cumbre, no solo con el rendimiento máximo, sino con el mejor desarrollo personal.

El deporte de alta competición se caracteriza por la exigencia de dos grandes procesos que los especialistas han denominado selección deportiva y entrenamiento deportivo. A la selección deportiva corresponde la determinación de aquellos individuos cuyas dotes tienen un carácter excepcional en las poblaciones. Al entrenamiento deportivo corresponde todo el proceso de desarrollo de las capacidades que determinan el rendimiento competitivo.

Resulta en extremo difícil la selección deportiva atendiendo al componente psicológico. Por ello, elegir un atleta con características psicológicas ideales resulta prácticamente imposible si no se ha tenido la oportunidad de apreciar su comportamiento psíquico en la actividad específica del deporte.

En el voleibol, el componente psicológico resulta para los entrenadores uno de los más difíciles en la planificación, ejecución y control del entrenamiento, ya que es complejo determinar las características psicológicas individuales, dada su variación por la diversidad de contextos en que se pueden encontrar los jugadores.

A pesar de ello, se observa que el aspecto menos trabajado dentro del componente psicológico es el proceso de análisis de las informaciones recolectadas de entrenamientos y competencias, y su empleo inmediato por el entrenador para la toma de decisiones. Por otra parte, es común el empleo de la estadística tradicional en el deporte, pero la misma empieza a tornarse inviable cuando la cantidad de variables e información a ser estudiadas aumenta.

Aunque se han estudiado diferentes aspectos psicológicos que intervienen en las acciones ofensivas y defensivas del voleibol, estas aún presentan insuficiencias a la hora de determinar cuáles se deben tener en cuenta.³

El entrenador requiere por ello un recurso que le permita de forma sintetizada lograr la información inmediata de las

características psicológicas del atleta para indicar las tareas de intervención psicológica que le correspondan.

Por tales motivos, el Centro de Estudios de Ingeniería de Sistemas (CEIS) del Instituto Superior Politécnico José Antonio Echeverría, Cujae, junto a la Federación Cubana de Voleibol, realizaron un estudio experimental para evaluar las posibilidades que ofrece la minería de datos en esta esfera.

A continuación se exponen las consideraciones tenidas en cuenta en ese estudio experimental y los resultados obtenidos al aplicar las técnicas de minería de datos.

CONSIDERACIONES DEL ESTUDIO EXPERIMENTAL

Para la realización del estudio experimental, la Federación Cubana de Voleibol, conjuntamente con el CEIS, consideraron conveniente aplicarlo a atletas juveniles participantes en las competencias nacionales, dada la importancia que le ofrece este país al deporte de base.

Con tales presupuestos, este estudio fue realizado con atletas del sexo masculino en la categoría de 15 a 16 años que intervinieron en los Juegos Escolares Nacionales celebrados en Santiago de Cuba del 9 al 19 de julio del 2002. De los 12 equipos participantes, se muestrearon los 8 mejores, según el ranking de los juegos del 2001. La muestra abarcaba 77 atletas.

Para la realización de este estudio se evaluó un conjunto de tests ampliamente conocidos y validados nacional e internacionalmente. Estos tests son conocidos con los siguientes nombres:³

- Mezcla temperamental.
- Rasgo de ansiedad.
- HAD.
- Motivos deportivos de Butt.
- Actitud en la competencia.
- Atención TAIS.
- Estado de ansiedad.
- Síntoma de carga.

El test **Mezcla temperamental** parte del principio de que todo sujeto posee en diferentes proporciones rasgos del temperamento, tales como sanguíneo, colérico, flemático y melancólico.

El test **Rasgo de ansiedad** mide el nivel de ansiedad que las personas tienen en su estado natural.

El test **HAD** mide el nivel de ansiedad y depresión que las personas tienen en su estado natural.

El test **Motivos deportivos de Butt (MD)** permite conocer el nivel de motivación del atleta. Este test valora las siguientes áreas de la motivación: conflicto, rivalidad, suficiencia, cooperación y agresividad.

El test **Actitud en la competencia** permite conocer la actitud del atleta. Este test valora las siguientes áreas: certeza de la fuerza, valoración de la fuerza, significado de la competencia y opinión de otras personas.

El **Test de atención TAIS** mide el nivel de atención de los atletas.

El test **Estado de ansiedad** mide el nivel de ansiedad que los deportistas tienen en estado de tensión psíquica (por ejemplo, antes de un juego).

El test **Síntoma de carga** mide el nivel de tolerancia psíquica del deportista en los aspectos: estabilidad neurovegetativa y de éxito-fracaso, estabilidad social-personal, y estabilidad ante situaciones objetivas del medio.

Además, la variable **rendimiento** fue considerada a partir de la teoría de las situaciones simplificadas de juego, la cual plantea que una situación en el juego se concibe como las acciones coordinadas de los integrantes del equipo desde el comienzo de la jugada hasta su conclusión.³

Con tal consideración, el rendimiento del atleta se verá expresado no como un desarrollo individual, sino como un comportamiento heterogéneo que contribuye al rendimiento del equipo en la competencia; por lo que quedó definido como: el porcentaje de sets ganados respecto al total de sets jugados en la fase eliminatoria y en la final. Esta variable **rendimiento** fue categorizada en 5 clases.

Considerando todas las variables presentes en los tests aplicados, junto a la variable **rendimiento**, se pudo observar un total de 25 variables, lo cual produce (para los 77 atletas) una tabla de 1925 valores; lo que, sin ser grande, representaba una magnitud intratable por medios manuales.

Para la realización de la minería de datos se evaluaron diferentes sistemas que, de forma libre, se podían obtener a través de Internet, tales como: See5, Tiberius, Weka, WizRule, entre otros. Tras un análisis de las ventajas y limitaciones de cada uno de estos sistemas se decidió utilizar el sistema Weka.

El sistema Weka es el resultado de una iniciativa de Ian H. Witten y Eibe Frank, ambos profesores de la Universidad de Waikato en Nueva Zelanda, con el propósito de implementar un paquete de minería de datos de código abierto (código fuente gratuito).⁵ El nombre de Weka proviene de Waikato Environment for Knowledge Analysis.

El sistema Weka está implementado en Java en su versión 3.3.6. Este sistema permite realizar tanto análisis descriptivos como predictivos. Para el análisis predictivo se tienen varias clases que implementan técnicas de clasificación, clustering y asociación. Para cada una de estas técnicas existen varios algoritmos, entre los que se tienen los siguientes:

- Clustering (weka.clusterers): EM, Simple KMeans, CobWeb, Farthest First y XMeans.

- Clasificación (weka.classifiers): J48 (C4.5), OneR, PRISM, ZeroR, Random Forest, REPTree, UserClassifier, HiperPipes, Red Neuronal del tipo Backpropagation, Lineal Regretion, Naive Bayes, y otros.

- Asociación (Weka. Associations): APriori.

Tras una evaluación de las facilidades que ofrece el sistema Weka, y de los algoritmos disponibles, se decidió por la realización de un análisis descriptivo de los datos recolectados, y por la aplicación de los algoritmos:

- EM: Por ser un método de clustering aplicable sobre datos categóricos.⁶

- J48: Por ser un método que genera árboles de clasificación aplicable sobre datos categóricos.⁷

- Apriori: Por ser un método que permite descubrir las asociaciones que existen entre categorías (ítems según el autor).

RESULTADOS OBTENIDOS

A continuación se explica brevemente el análisis descriptivo realizado, así como los algoritmos utilizados y los resultados obtenidos sobre los datos de la fase eliminatoria.

Análisis descriptivo

El análisis descriptivo brinda información del estado actual de la base de datos, llegando a un conocimiento de nivel primario.

El sistema Weka incluye varios recursos para el análisis descriptivo, destacándose las facilidades de visualización. Con estos recursos pudo apreciarse, en los datos de la fase eliminatoria, diferencias considerables en algunas variables psicológicas para los atletas de muy alto rendimiento. A continuación se muestra la salida del sistema con algunas de estas variables, siendo el gráfico inferior derecho (**rendimiento**) la leyenda del resto.

En la figura 1 puede observarse que los atletas de rendimiento muy alto (en color rosado) presentan un comportamiento más homogéneo, verificándose para estos atletas, entre otros, que:

- No presentan **ansiedad (HAD)** alta (gráfico superior izquierdo).

- Presentan una **depresión (HAD)** y **agresividad (MD)** media de forma significativa (gráficos superior medios).

- No presentan temperamentos **coléricos**, muy bajo, bajo, pero tampoco muy alto (gráfico inferior izquierdo).

- No presentan **rasgos de ansiedad** bajo (gráfico inferior medio).

Algoritmo EM

Es un algoritmo de clustering que calcula la probabilidad de un cluster a partir de determinados parámetros de distribución del conjunto de datos. Este algoritmo define puntos (semillas) dispersos aleatoriamente, generando de forma iterativa las mejores distribuciones probabilísticas entre los clusters.⁹⁻¹¹

Al aplicar el algoritmo EM, el sistema Weka ofreció los

```

Class attribute: Rendimiento
Classes to Clusters:

0  1  2  3  <--- assigned to cluster
0  0  0  0  | Muy Bajo
4  5  1  0  | Bajo
2  7  0  0  | Medio
4  31 3  1  | Alto
9  8  1  1  | Muy Alto

Cluster 0 <--- Muy Alto
Cluster 1 <--- Alto
Cluster 2 <--- Bajo
Cluster 3 <--- No class
    
```

siguientes resultados:

Para comprender mejor estos resultados puede observarse el gráfico de la figura 2.

A partir de estos resultados pudieron comprobarse diversos comportamientos, entre los que se tienen, a modo de ejemplo, los siguientes:

- El **rendimiento medio** no fue caracterizado como una clase homogénea independiente, agrupándose con los de la clase: Alto y Muy Alto.
- Los atletas de **rendimiento alto** representan un grupo muy homogéneo (el 79,4%; o sea, 31 de 39 atletas) respecto al comportamiento psicológico.
- La clase de **rendimiento muy alto**, dentro del mismo cluster, queda muy pobremente representada (47,3%; o sea, 9 de 19

atletas).

Algoritmo J48

El algoritmo J48 constituye una implementación del ID3, más específicamente del C4.5. El ID3 es considerado un clásico de los algoritmos de aprendizaje automático, el cual construye árboles de decisión a partir de conjuntos de ejemplos, de allí su nombre (Induction of Decision Trees). Este algoritmo fue desarrollado por Quinlan en la primera mitad de la década del 80.⁹

El algoritmo C4.5 (o J48), al igual que ID3, genera árboles de decisión. Sin embargo, en él se destacan los siguientes aspectos: la poda (simplificación del árbol de decisión), la manipulación de valores de atributos omitidos, el uso de la

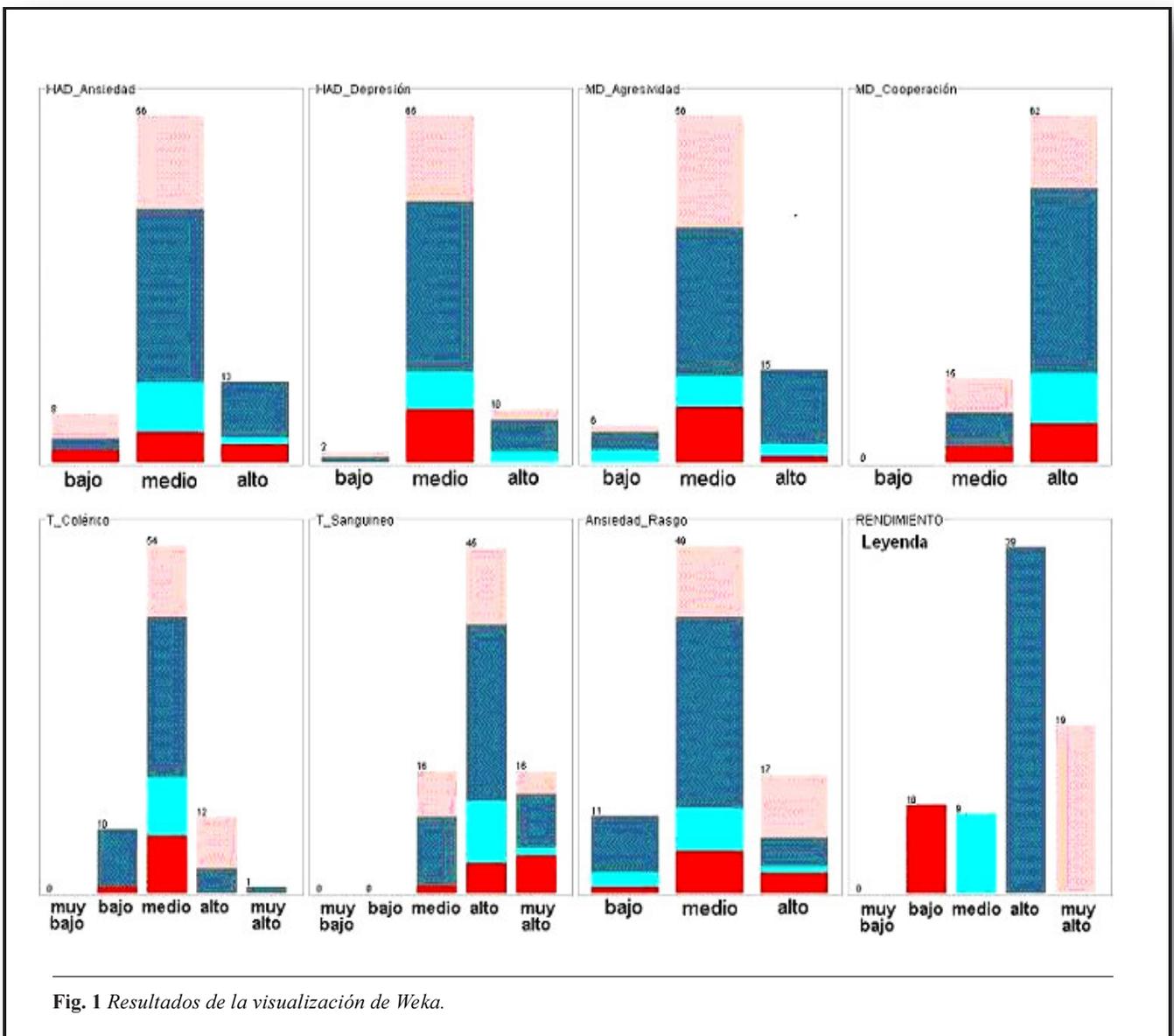


Fig. 1 Resultados de la visualización de Weka.

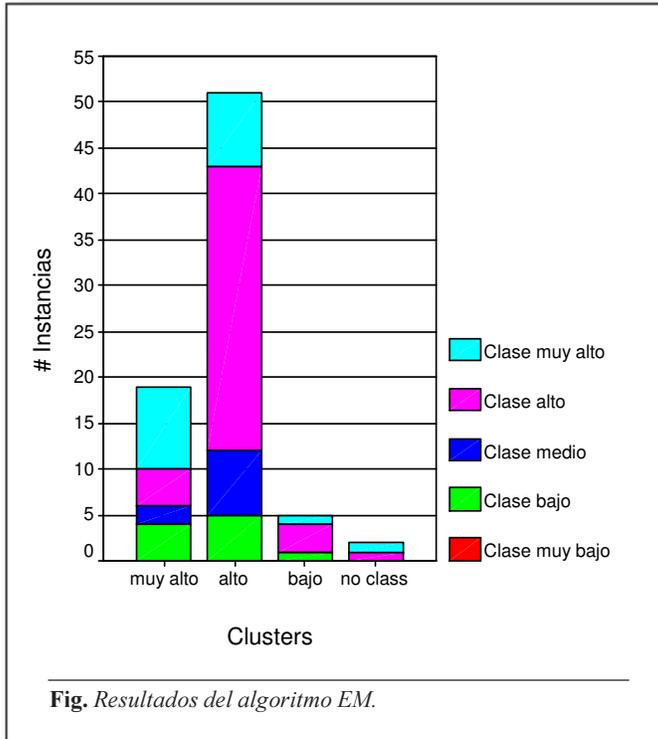


Fig. Resultados del algoritmo EM.

desigualdad, y los conjuntos para tratar atributos numéricos.⁷

Al aplicar el algoritmo J48, se obtuvo un árbol de decisión el cual se muestra parcialmente a continuación:

```

T_Colérico = Bajo: Alto (10.0/1.0)
...
T_Colérico = Alto
| Agresividad = Bajo: Alto (1.0)
| Agresividad = Medio
|| Ansiedad = Bajo: Muy Alto (2.0)
|| Ansiedad = promedio: Muy Alto (6.0)
|| Ansiedad = Alto: Alto (1.0)
| Agresividad = Alto: Alto (2.0)
T_Colérico = Muy Alto: Alto (1.0)
    
```

Number of Leaves: 37
Size of the tree: 52

A partir de estos resultados pudieron comprobarse diversos comportamientos, entre los que se tienen, a modo de ejemplo, los siguientes:

- Seis atletas con el **temperamento colérico alto**, la **agresividad media** y la **ansiedad no alta** tuvieron **muy alto rendimiento**.
- Diez atletas con el **temperamento colérico bajo** tuvieron **alto rendimiento**.
- La ausencia de la variable **cooperación** en todas las ramas del árbol de la fase eliminatoria (no así en el árbol de la fase final).

Algoritmo APriori

Este algoritmo fue propuesto por R. Agrawal y R. Srikant para crear reglas de asociación.⁸ El nombre del algoritmo está dado

por el uso del "conocimiento previo" de las frecuencias del conjunto de instancias. El algoritmo APriori emplea una aproximación iterativa conocida como búsqueda de "nivel sabio", donde conjuntos de k instancias son usados para explorar conjuntos de k+1 instancias.

Al aplicar el algoritmo APriori se obtuvieron (en la fase eliminatoria) 15 reglas, todas con la variable **cooperación** presente como consecuente de las reglas. Ejemplo de las reglas de asociación obtenidas es el siguiente:

El 97 % de los atletas con la **valoración del contrario alto** y la **suficiencia alta** y el **rasgo de ansiedad medio** también mostraron la **cooperación alta**, verificándose esto en el 38,9% de los casos (30 de 77).

Además de los resultados que cada una de estas técnicas ofrece, se mostró la conveniencia de evaluar la interrelación del conocimiento descubierto entre todas ellas, lo cual permite analizar las variables que inciden en el alto o bajo rendimiento en este deporte.

En particular, el análisis descriptivo ofrece una comprensión inicial de las variables con una alta significación respecto al rendimiento o, incluso, las variables que se esperaban tuvieran una significación apreciable y no la tienen. Por otra parte, el algoritmo EM permite conocer de forma preliminar el grado de similitud de los atletas con igual rendimiento.

Teniendo tal conocimiento previo sobre los factores psicológicos estudiados, la aplicación del algoritmo J48 facilita la caracterización de los atletas con comportamientos similares a partir de la conjunción de factores que determinan de forma significativa los rendimientos requeridos. Además, sobre las variables identificadas, tanto las que inciden directamente en el alto o bajo rendimiento como las que no inciden pero de las que se esperaba otro comportamiento, pueden evaluarse las asociaciones o relaciones que muestran entre sí, o con otras variables, al aplicar el algoritmo APriori.

Aunque los resultados de este estudio no pueden extrapolarse mecánicamente a los siguientes juegos escolares, lo cual requeriría de datos de varios años y efectuar análisis de secuencias, este descubrió determinados comportamientos que resultaron de gran interés para los especialistas deportivos. Estas regularidades permitieron la planificación de estudios más específicos y la conformación de programas que permitirán modificar o madurar algunos de los rasgos que inciden en el rendimiento.

CONCLUSIONES

Como ha podido apreciarse, la minería de datos ofrece grandes posibilidades para el análisis de los datos, permitiendo el descubrimiento de patrones y regularidades de alta significación. A pesar de que la minería de datos ha sido aplicada fundamentalmente en el marco empresarial, su potencialidad en la esfera deportiva es promisoría.

El estudio experimental realizado sobre el rendimiento psicológico en el voleibol confirmó que la minería de datos puede ayudar a entrenadores y especialistas a tomar mejores decisiones, posibilitando la obtención de informaciones y tomar decisiones a las que normalmente no se podía llegar.

En este estudio pudo apreciarse, en el análisis descriptivo, los comportamientos de algunas variables psicológicas, descubriendo nuevas variables significativas o, incluso, descartando otras tradicionalmente consideradas como tal para el rendimiento de los atletas.

Con el clustering pudo evaluarse el grado de semejanza entre los atletas con iguales rendimientos. Mientras tanto, en la clasificación, se seleccionaron las variables más significativas para decidir el rendimiento. Además, con la generación de las reglas de asociación significativas, pudieron evaluarse las asociaciones o relaciones que muestran las variables identificadas anteriormente.

En general, el estudio realizado permitió la evaluación de elementos metodológicos que pueden orientar a los expertos del deporte en la búsqueda e investigación de determinados factores que inciden en el rendimiento de los atletas, mostrando la posibilidad de trabajar en esta línea.

Como resultado de este estudio experimental, los especialistas y directivos deportivos han comprendido la necesidad y conveniencia de aplicar estos resultados, promoviendo un estudio más profundo en esta área. ☐

REFERENCIAS

1. REESE, S.: "Searching for the Mother Lode: Tales of the First Data Miners. Intelligent Systems & their Applications", *IEEE Expert*, Vol. 11, No. 5, pp. 4-7, Oct. 1996.
2. FIELITZ, L. AND D. SCOTT: *Prediction of Physical Performance Using Data Mining*, The 2003 AAHPERD National Convention and Exposition, 2003.
3. ANDUX, C. Y E. GEORGE: "Nuevo enfoque para el aprendizaje de los juegos deportivos: El voleibol, *Rev. del Instituto Superior de Cultura Física Manuel Fajardo*, Ciudad de La Habana, Cuba, 2000.
4. PINHEIRO, M.: "La minería de datos en función del diagnóstico psicológico del rendimiento de los voleibolistas", Tesis de maestría, CEIS, Cujae, Ciudad de La Habana, 2003.
5. <http://www.cs.waikato.ac.nz/ml/weka/index.html>.
6. HASTIE, TREVOR; ROBERT TIBSHIRANI AND JEROME FRIEDMAN: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer-Verlag, 2001.
7. QUINLAN, R.: *C4.5: Programs for Machine Learning*. Morgan Kaufman Publishers, Inc., 1993.
8. AGRAWAL, R. AND R. SRIKANT: *Fast Algorithms for Mining Association Rules in Large Databases*. Proc International Conference on Very Large Databases, pp. 478-499, Morgan Kaufmann, 1994.
9. MITCHELL, T.: *Machine Learning*: WCB/McGraw-Hill, 1997.
10. WITTEN, I. AND E. FRANK: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2000.
11. HAN, J. AND M. KAMBER: *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2001.

REVISTAS CIENTÍFICAS DE LA CUJAE AHORA EN FORMATO ELECTRÓNICO ¡VISÍTENOS!



• <http://intranet/ediciones/>