

# USO DEL SOFTWARE CURRICULAR COMO TÉCNICA DE APRENDIZAJE SUPERVISADO POR COMPUTADORAS

***Autores:*** Gerardo González Macías<sup>1</sup>, Humberto Chaviano Arteaga<sup>2</sup>, Víctor Yambay Delgado<sup>3</sup>

***Correo Electrónico:*** [gggm\\_jerry@yahoo.com](mailto:gggm_jerry@yahoo.com)<sup>1</sup>

***Institución:*** Instituto Superior Tecnológico Bolivariano de Tecnología, Ecuador<sup>1, 2 y 3</sup>

## INTRODUCCIÓN

Durante miles de años, el hombre ha tratado de entender cómo piensa y razona a corto o a largo plazo. Esta comprensión ha sido vital y ha marcado la diferencia en el desarrollo de la sociedad tal como se conoce hasta hoy. Este conocimiento también se ha combinado con el poder de las computadoras, tratando de simular o lograr un “aprendizaje” en éstas; se intenta unir de esta manera lo mejor de ambos mundos: la agilidad del cómputo de las máquinas y la inteligencia humana. A través del tiempo la misma se ha visto enriquecida con los aportes de Sistemas Basados en el Conocimiento (1969-1979) en el desarrollo de mecanismos de búsquedas de propósito general, para encontrar soluciones completas y, posteriormente, convertida en una Ciencia (desde 1987 hasta el presente) dando surgimiento a los modelos conexionistas conocidos como los Árboles de Decisión que aplican un algoritmo de aprendizaje de retroalimentación aplicados en los campos de la informática y la pedagogía. Actualmente el uso del software curricular ha sido una herramienta pedagógica muy fundamental para difundir en los estudiantes de sistemas la práctica de cómo enseñar a que las computadoras piensen como seres humanos. El presente trabajo tiene como objetivo dar a conocer un tipo de Aprendizaje muy utilizado en la especialidad de Tecnología en Análisis de Sistemas Computacionales mediante el uso de la herramienta informática WEKA: el Aprendizaje Supervisado. Se explicarán los Árboles de

Decisión y la importancia de su uso en el contexto del uso de los softwares curriculares.

## **DESARROLLO**

De acuerdo con (Michalski, 1986), aprender es la habilidad de adquirir nuevo conocimiento, desarrollar habilidades para analizar y evaluar problemas a través de métodos y técnicas, así como también por medio de la experiencia propia; siendo un requisito que el resultado del aprendizaje sea entendible para un hombre.

En el año 1956, durante un Congreso en Dartmouth (E.U) se propuso este término para agrupar a todos los métodos, técnicas e intentos de simular el intelecto humano en la computadora. Se determinó que un “programa inteligente” es uno que muestra un comportamiento similar al humano cuando se enfrenta a un problema. No es necesario que el programa resuelva realmente el problema de la misma forma que el hombre. De igual forma, un gran investigador del aprendizaje de computadoras, Alan Turing expresó: “si durante el intercambio entre una computadora y el usuario este último cree que está intercambiando con otro humano, entonces se dice que el programa es inteligente”. Esta idea queda reflejada en la imagen que muestra el anexo 1. Sobre la base de estos principios sobre el Aprendizaje y las aspiraciones por lograr que la computadora interactúe con el hombre como otro humano más, se define Aprendizaje Automatizado enfocado en el diseño y desarrollo de algoritmos y técnicas que permitan a las computadoras "aprender". Como herramienta computacional y pedagógica, en el ITB se usan los softwares curriculares. Se define el concepto de Software Curricular: es una aplicación informática, que se sustenta sobre una bien definida estrategia pedagógica, responde a los objetivos de un programa y al proyecto educativo de la sociedad y de la institución en la que se utilice (Labañino R. César. 2001). El software curricular contribuye, además, a elevar el rol protagónico del estudiante, pues le permite seleccionar el contenido a estudiar aprovechando la estrategia instructiva especificada por el autor, para conducir el

proceso de aprendizaje. Sobre la base de que a las máquinas hay que indicarles cómo aprender, se define el concepto de Aprendizaje Supervisado, el cual es posible realizarlo mediante el uso de un software apropiado. El algoritmo produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema. Constituye un algoritmo de aprendizaje basado en ejemplos donde el nuevo conocimiento es inducido a partir de una serie de ejemplos y contraejemplos. Un ejemplo de este tipo de algoritmo es el problema de clasificación, donde el sistema de aprendizaje trata de etiquetar (clasificar) una serie de vectores utilizando una entre varias categorías (clases). La base de conocimiento del sistema está formada por ejemplos de etiquetados anteriores. Entre las técnicas de Aprendizaje Supervisado tenemos: los árboles de decisión.

**Árbol de Decisión (Decision Trees o DT):** son una forma de representación sencilla, muy usada entre los sistemas de Aprendizaje Supervisado, para clasificar ejemplos en un número finito de clases. Se basan en la partición del conjunto de ejemplos según ciertas condiciones que se aplican a los valores de los atributos (Se sugiere observar la imagen mostrada en el anexo 2). Los sistemas basados en DT forman una familia llamada TDIDT (Top-Down Induction of Decision Trees), cuyo representante más conocido es el algoritmo ID3<sup>1</sup>

## PROBLEMA

La base de casos está relacionada con patrones que permitieron realizar un reconocimiento de tres tipos de vinos (“wine.arff”). Estos datos son los resultados de un análisis químico realizado a distintas clases de vinos cosechados en la región de Via Brigata Salerno, Italia. Dicha base fue construida a partir de una serie de datos coleccionados por el “*Institute of Pharmaceutical and Food Analysis and Technologies, Genoa, Italy*”. En ella están presentes 13 variables ( $X_1, X_2, \dots, X_{13}$ ) que constituyen características o rasgos encontrados en cada uno de los vinos y una variable  $X_{14}$  que representa la “clase  $i$ ” ( $i=1, 2, 3$ ) con sus

---

<sup>1</sup> ID3 (Interactive Dichotomizer) se basa en la reducción de la entropía media para seleccionar el atributo que genera cada partición (cada nodo del árbol), seleccionando aquél con el que la reducción es máxima. Los nodos del árbol están etiquetados con nombres de atributos, las ramas con los posibles valores del atributo, y las hojas con las diferentes clases.

correspondientes instancias. Se pretende obtener un Modelo Matemático que permita clasificar cada clase de vinos (Tipo a, Tipo b, Tipo c). Se conoce previamente que los clasificadores no se pueden generalizar, es decir, “no existe el clasificador perfecto de todos los clasificadores”; por lo que se hará énfasis sobre el modelo de clasificación de los Árboles de Decisión, como herramienta de aprendizaje supervisado y se interpretarán los resultados obtenidos haciendo uso de los algoritmos ID3 y J48. De manera general, la base de casos contiene las siguientes características:

Número de Instancias: 178

Área: Química

Distribución de clases: número de instancias por clases

*Clase 1: 59*

*Clase 2: 71*

*Clase 3: 48*

Número de Atributos: 13

Características de los Atributos: Continuos

Fecha de Confección: Julio 2010

## **USO DEL WEKA COMO SOFTWARE CURRICULAR**

Las técnicas de análisis se realizarán en el software curricular Weka (en su versión 3.4-2). La licencia de Weka es GPL<sup>2</sup> (**GNU Public License**), lo que significa que este programa es de libre distribución y difusión. Además, dado que Weka está programado en Java, este es independiente de la arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual de Java **disponible (Sugerencia: Ver Anexo 3). Se recomienda además ver el anexo 4** que muestra una imagen con la información de las 14 variables que intervienen en el problema, contenidas en la base de datos “wine.arff”. Como herramienta pedagógica e informática el estudiante de sistemas hará uso de los Árboles de Decisión con el algoritmo computacional J48 para atributos continuos. Observe el

---

<sup>2</sup> <http://www.gnu.org/copyleft/gpl.html>

clasificador que se construye con el WEKA: *Classify/Choose/J48-C 0.25 –M*  
 2.Una vez ejecutada la acción, se despliega la siguiente información:

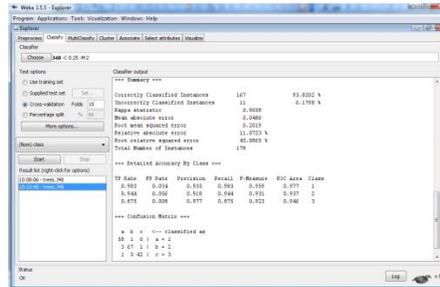


Fig. 2: Resultados Matemáticos del aprendizaje realizado por la computadora con la función J48. Fuente: Elaborado por Gerardo González.

El estudiante ahora debe ser capaz de interpretar la información matemática y emitir un pequeño informe, el cual más o menos debe contener lo siguiente:

## MATRIZ DE CLASIFICACIÓN

El 93.8202 % de los casos correctamente clasificados, lo que muestra un tamaño de 167 instancias. El 6.1798 % que corresponde a las 11 restantes, fueron mal clasificados. La matriz de confusión muestra en su primera fila que de 59 casos de la clase  $a=1$ , se equivocó en 1 de ellos identificándolos como clase "b". A su vez, la segunda fila muestra que de 71 casos de la clase  $b=2$ , se equivocó en 4 de ellos, identificando 3 como clase "a", y uno como "c". La última, por su parte, muestra que de 48 casos de la clase  $c=3$ , clasificó mal 6 de ellos, asociando uno de ellos con la clase "a" y 5 con la "b". En total se obtuvieron 11 casos mal clasificados. Para generar el Árbol de Decisión construido por J48, se hace clic en la opción "Visualize Tree" (Sugerencia: Ver Anexo 5).

## DESCRIPCIÓN DEL ÁRBOL DE DECISIÓN GENERADO

Como es de suponer, el nodo raíz representa el rasgo que mejor discrimina ("Mayor Ganancia"), por tanto, a partir de  $x_7$  comienza la primera subdivisión: para valores de  $x_7 \leq 1.57$  el nodo candidato a expandirse en el subárbol izquierdo es el que contiene el rasgo  $x_{10}$ ; para valores de  $x_{10} \leq 3.8$ , se clasifican correctamente 13 ejemplares de vinos (con estas características) como clase  $b=2$ , sin embargo para valores de  $x_{10} > 3.8$ , de los 50 ejemplares (con

estas características), se clasificaron correctamente sólo 49, es decir, hubo un ejemplar que lo identificó como clase  $c=3$  cuando en realidad no lo es. La interpretación para  $x_{13}$  y  $x_{10}$  en el subárbol derecho es análoga. A partir del árbol se pueden derivar un conjunto de reglas de decisión que ayudan a cualquier especialista, no necesariamente experimentado en computación, a entender cuándo un vino es de una clase u otra. Por ejemplo:

**Si  $x \leq 1.57$  y  $x_{10} \leq 3.8$  entonces el vino es de clase 2      CERT (Regla)=1**

## **CONCLUSIONES**

De acuerdo con lo investigado, se obtienen las siguientes conclusiones:

Fue posible realizar una breve descripción de los conceptos de Aprendizaje Supervisado, Software Curricular y Árbol de Decisión, estableciéndose una relación entre cada uno de ellos, mediante la solución computacional de un problema de la vida real. Durante la ejecución de la solución, el estudiante fue capaz de determinar una solución mediante la clasificación de los árboles de decisión, donde generó un modelo matemático en el cual el 93.8202 % de los casos de vinos fueron correctamente clasificados, obteniendo un árbol de tres niveles, cuyo nodo final representaba a tres clases de vinos por separado.

## **RECOMENDACIONES**

- Profundizar con el estudiante, en los conocimientos prácticos del Aprendizaje Supervisado de Computadoras que realizan las Redes Neuronales Artificiales (RNA) en el área de clasificación de conceptos, usando el WEKA como herramienta computacional y pedagógica.
- Lograr que el estudiante sea capaz de mostrar resultados prácticos del aprendizaje de las RNA tomando como punto de partida el mismo problema de clasificación de los vinos con la base de datos "wine.arff".

## BIBLIOGRAFÍA

- Lerner, I. YA. Fundamentación didáctica de los métodos de enseñanza. (Traducción del Ruso). Hemeroteca del MINED, 1986.
- Labañino Rizzo, Cesar. Multimedia para la educación. Editorial Pueblo y Educación, la Habana. 2001.
- MICHALSKI, R. S. 1986. "Understanding the Nature of Learning: Issues and Research Directions". In: R. S. MICHALSKI, J. G. C. Y. T. M. M. (ed.) *Machine Learning: An Artificial Intelligence Intelligence Approach*. Los Altos, California: Morgan Kaufmann Publishers.
- MICHIE, D. 1991. "Methodologies for machine learning in data analysis and software". *Computer Journal*.
- NORVIG, S. J. R. Y. P. 2004. "Historia de la Inteligencia Artificial". In: ARAGÚN, D. F. (ed.). *Inteligencia Artificial. Un Enfoque Moderno*. Segunda Edición ed. Madrid, Spain: PEARSON EDUCACIÓN, S.A.
- QUINLAN, J. R. 1986. "Induction of Decision Trees". *Machine Learning*.
- QUINLAN, J. R. 1990. "Learning Logical Definitions from Relations". *Machine Learning*.
- QUINLAN, J. R. 1993. "C4.5: Programs for Machine Learning".
- QUINLAN, J. R. 1996. Learning first-order definitions of functions. *Journal of Artificial Intelligence Research*, 5, 139-161.

## ANEXOS

### Anexo 1: Esquema relacionado con la Interacción Estudiante- Computadora

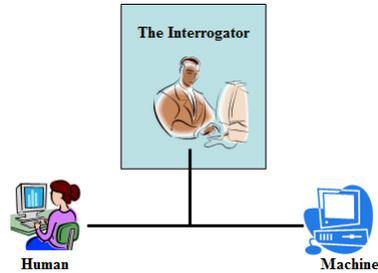


Fig. 1.1: Interacción Estudiante-Computadora. Fuente: Elaborado por Gerardo González

### Anexo 2: Vista computacional de un árbol de Decisión

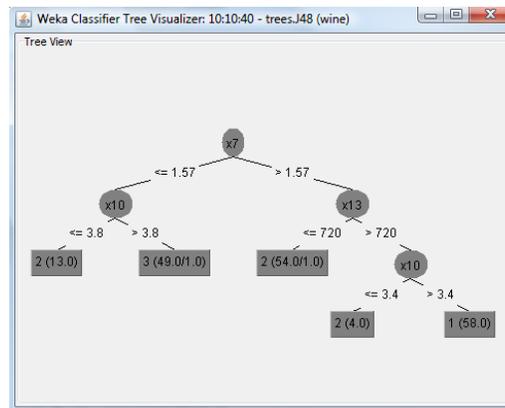


Fig. 1.2: Árbol de Decisión de 3 niveles. Fuente: Elaborado por Gerardo González

### Anexo 3: Ambiente de Trabajo del Software Weka

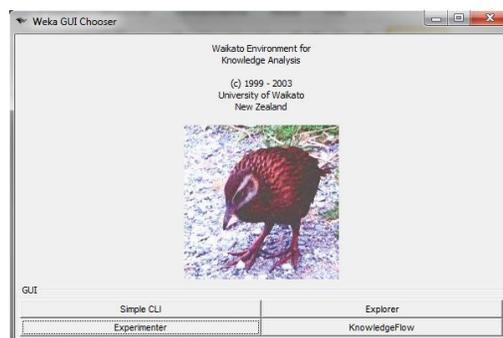


Fig. 1.3: Ambiente de Trabajo del Software Weka. Fuente: Universidad de Waikato, Nueva Zelanda.

#### Anexo 4: Información de la base de datos “wine.arff” leída desde Weka

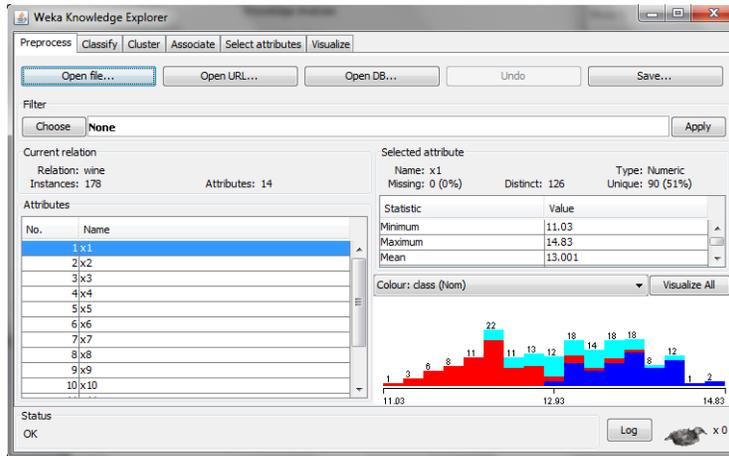


Fig. 1.4 Información de la base de datos “wine.arff” leída desde Weka. Fuente: Elaborado por Gerardo González

#### Anexo 5: Árbol de Decisión de tres niveles generado en la solución del problema

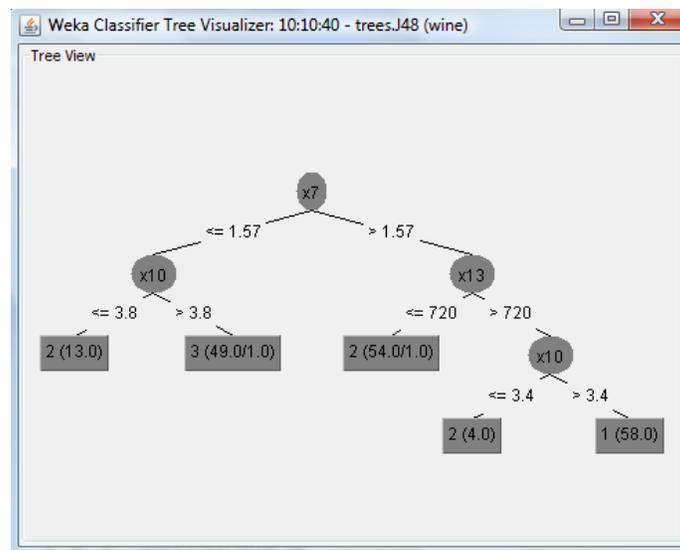


Fig. 1.5: Árbol de Decisión de tres niveles. Fuente: Elaborado por Gerardo González.