



**XIX COLÓQUIO INTERNACIONAL DE GESTÃO UNIVERSITÁRIA**  
*Universidade e Desenvolvimento Sustentável: desempenho acadêmico e os desafios da sociedade contemporânea*

Florianópolis | Santa Catarina | Brasil  
25, 26 e 27 de novembro de 2019  
ISBN: 978-85-68618-07-3



## **ALGORITMOS DE DECISIÓN APLICADOS AL RENDIMIENTO ACADÉMICO**

**Santiago Noguera Valdez<sup>1</sup>**

Universidad Columbia del Paraguay  
[santiagoanoguera@gmail.com](mailto:santiagoanoguera@gmail.com)

**Roberto Sánchez<sup>2</sup>**

Universidad Columbia del Paraguay  
[totosanchez@gmail.com](mailto:totosanchez@gmail.com)

**Heriberto Pintos Correa<sup>2</sup>**

Universidad Columbia del Paraguay  
[Heriberto.pintos@columbia.edu.py](mailto:Heriberto.pintos@columbia.edu.py)

### **RESUMEN**

El apoyo a la gestión temprana del proceso enseñanza-aprendizaje genera una ventaja importante en dicho proceso. En tal sentido, este trabajo propone una aplicación que permite evaluar al alumno en base a atributos, para determinar la probabilidad de finalización de la materia con éxito. El objetivo de la aplicación es dotar al docente de una herramienta tecnológica capaz de analizar el grado de avance del alumno y su probabilidad de éxito en una determinada materia, teniendo en cuenta los registros del sistema. Los atributos que forman parte de la evaluación son procesados por un algoritmo y las variables seleccionadas soportadas por la fórmula de entropía. El diseño de la solución es web responsivo.

**Palabras Clave:** Algoritmo de decisión, proceso enseñanza-aprendizaje, diseño web responsivo, fórmula de entropía.

<sup>1</sup> Ing. en Informática,

<sup>2</sup> Docente

## **Introducción**

La labor docente en clase se vuelve cada día más compleja teniendo en cuenta su rol fundamental de transferencia de conocimientos a los alumnos. El docente debe tratar de desarrollar contenidos y que éstos sean asimilados para el éxito en la materia. La diversidad de niveles de conocimiento entre alumnos es una variable considerable. Debido a esto, brindar al docente la posibilidad de evaluar a los alumnos en etapas tempranas de la materia podrá darle una perspectiva sobre la condición de los mismos. El docente juega un papel importante en el desarrollo de la materia, pero así también el poder aprender y aprobar la materia depende del alumno, quien por su parte puede estar influenciado por varios factores, como ser la asistencia a clases, el interés por la materia, el nivel económico, entre otros. Conocer cuáles son los factores que influyen en la probabilidad de aprobar la materia puede ser de gran ayuda al docente y beneficioso para el alumno. El apoyo a la gestión temprana del proceso enseñanza-aprendizaje genera una ventaja importante en dicho proceso. En tal sentido, se propone una aplicación que permita evaluar al alumno en base a atributos, para determinar la probabilidad de finalización de la materia con éxito.

Los resultados obtenidos en la presente investigación forman parte del trabajo de conclusión de carrera de ingeniería en informática de la Universidad Columbia del Paraguay (Noguera, 2018).

### **1. Objetivos y método**

Implementar una aplicación móvil utilizando árboles de decisión de apoyo al docente permite la evaluación previa al examen final. Para el logro de dicho objetivo se seleccionan las variables a ser utilizadas en el árbol de decisión por medio de la fórmula de entropía, se elabora el esquema de la comunicación entre la aplicación web/móvil y el servidor, la aplicación permite el registro y la evaluación del alumno.

La forma de investigación es aplicada y para los datos de prueba se ha seleccionado materias de cinco carreras, con datos ficticios para demostrar el funcionamiento de la aplicación.

### **2. Bases teóricas**

Para generar los árboles de decisión es preciso contar con datos de entrenamiento, los cuales servirán para clasificar y predecir el resultado buscado.

La predicción en la probabilidad de aprobar una materia se basa en ciertos atributos de los alumnos, los cuales se encuentran sustentados en un trabajo de investigación denominado: “Los docentes: funciones, roles, competencias necesarias, formación” (Marqués, 2000).

A medida que van siendo analizados nuevos registros, estos pueden formar parte del conjunto de datos de entrenamiento, es decir, si al principio se cuenta con 20 registros y a partir de ellos fueron analizados 15 alumnos nuevos, es posible incluir a estos para tener un total de 35 registros de entrenamiento.

Los árboles de decisión generados pueden ir variando de acuerdo a los datos ingresados, y para cada carrera estos árboles son distintos y cambiantes, pudiendo modificar constantemente su estructura.

La aplicación desarrollada es web responsiva, por lo que se podrá ver tanto desde una pc o notebook, como así también desde el celular. Al ser responsiva le permite ajustarse a la pantalla del dispositivo desde el cual es visualizado.

Se ha seleccionado cinco carreras de muestra para realizar las pruebas con el algoritmo de árbol de decisión. Los años de carrera se encuentran definidos del primer hasta el quinto año.

Para poder aplicar el algoritmo de árbol de decisión se precisa de una base de datos cargada con atributos de los alumnos.

En la base de datos se tienen cargados un total de 25 materias para las cinco carreras de muestra, los cuales se dividen por año basados en el planeamiento académico.

### **3.1. Antecedentes y teoría básica**

Cardona (2004), con un enfoque financiero utiliza árboles de decisión en el trabajo: “Aplicación de árboles de decisión en modelos de riesgo crediticio”, en dónde el algoritmo se utiliza como herramienta para el cálculo de probabilidades de incumplimiento en créditos.

En el trabajo de Valero et al. (s.f), “Predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos”, se

utilizan árboles de decisión C4.5 que son categorizados como aprendizaje basado en similitudes.

Otra aplicación está basada en la asistencia a clase y de apoyo docente (Barchino et al., 2001).

Según Rokach y Maimon (2015) “el árbol de decisión es un modelo de predicción que es utilizado en diversos ámbitos que van desde la inteligencia artificial hasta la economía”.

La entropía  $H(X)$  de una variable  $X$  se define por:

$$H(x) = - \sum p(x) \log p(x)$$

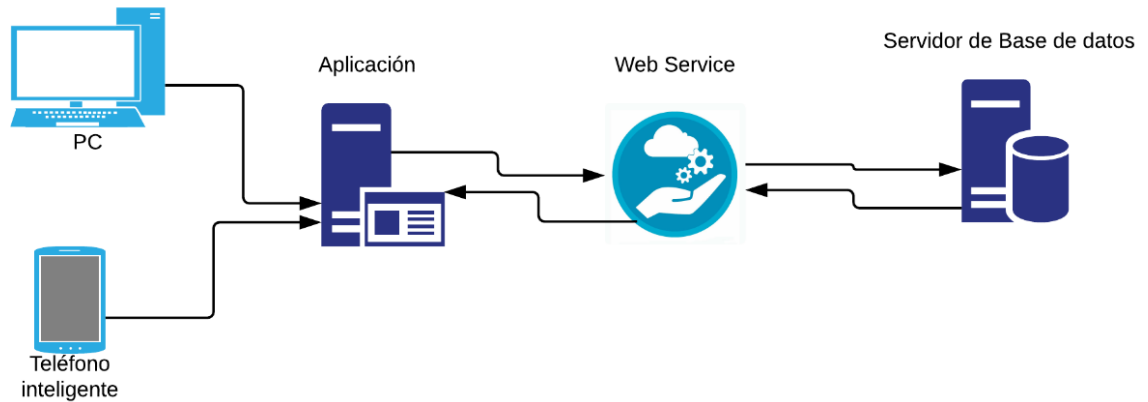
Por ejemplo, la entropía de un lanzamiento de moneda es de 1 bit. Se utiliza la convención de que  $0 \log 0 = 0$ , que se justifica fácilmente por la continuidad ya que  $x \log x = 0$  como  $x = 0$ . Por lo tanto, la adición de términos de probabilidad cero no cambia la entropía (Cover & Thomas, 1991).

Para Chesñevar (sf) la Ganancia de la Información es, “La heurística para elegir un test en un nodo, elegir el test que provee mayor información acerca de la clase. Este es el test que, en promedio, reduce más la entropía de la clase”. Como se puede ver en la formula, la reducción esperada de la entropía es igual a la ganancia de la información:  $Gain(S,A) = Entropía(S) - \sum_{v \in \{weak, strong\}} (|S_v|/|S|) * Entropía(S_v)$ , en donde:  $A =$  Atributo y  $S_v =$  Subconjunto de  $S$  para el cual el atributo  $A$  tiene valor  $v$ .

### 3. Arquitectura

El diseño de la aplicación es Web Responsiva y para el desarrollo se ha utilizado el IDE Netbeans (Fig. 1), permite la utilización de código java como base fundamental del proyecto, teniendo en cuenta la variedad de librerías gratuitas proveídas por esta herramienta y que permite una manera más flexible a la hora de realizar la programación.

**Figura 1.** Arquitectura del sistema.



#### 4. Diagrama de Entidad-Relación

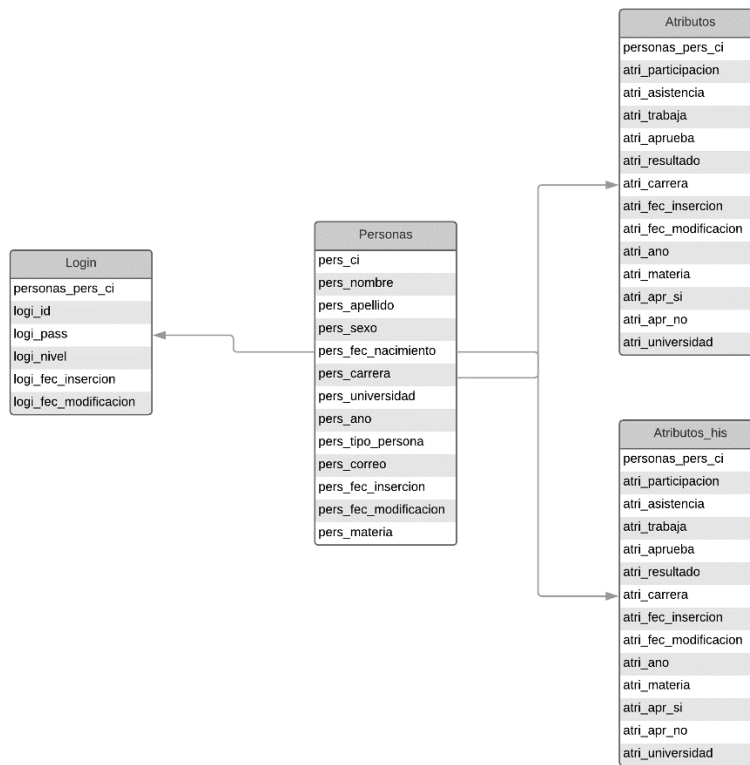
En el diagrama de entidad relación se definen cuatro tablas, creadas en Mysql. En la tabla de Personas se guardan los datos sensibles tanto de los alumnos como de los docentes, por ejemplo, el número de cédula de la persona que actúa como clave primaria, el nombre, apellido y fecha de nacimiento, entre otros (Fig.2).

En la tabla *Login* se guardan los datos de los usuarios ya registrados en la aplicación utilizando la clave principal y un campo de seguridad. Posee un campo de nivel, que indica que tipo de accesos tiene dentro de la aplicación, sirve para poder ingresar como administrador o como un simple usuario.

Para la tabla Atributos, se definieron columnas que guardan la información de los atributos de cada alumno, como por ejemplo su participación, su asistencia a clases, si trabaja o no. En esta tabla se registra si el alumno aprueba o no la materia y con qué porcentaje.

La tabla Atributos\_*his* es el histórico de la tabla Atributos, cuya función es dar cierto respaldo a los datos que se actualizan en la tabla con un historial por alumno de acuerdo a los atributos evaluados.

**Figura 2. Diagrama Entidad Relación de la aplicación.**



## 5. Reglas para la transformación de los datos

Para poder realizar la interacción con una base de datos de alumnos que no posea el formato con el cual se desea trabajar, se previeron algunas reglas de transformación y se realiza una demostración de inserción a la base con los datos transformados.

Al momento de cargar registros en la base de datos se tienen los siguientes atributos que definen el análisis de asistencia, trabajo y participación en clase. Para poder realizar una transformación de datos más precisa se pueden clasificar las materias en: mayormente prácticas o mayormente teóricas, ya que podrían existir materias que requieran de un mayor porcentaje de asistencia o participación que otras.

Se pueden definir las siguientes reglas: Asistencia a clase: si el campo es menor al 70% entonces, el dato transformado será “Deficiente”. Si es mayor o igual al 70% entonces será “Satisfactorio”.

Trabajo en clase: se toma el 70% del puntaje total acumulado, si es menor al 70% entonces no posee el mínimo para trabajos en clase, si es mayor o igual al 70% cumple con lo requerido.

Participación en clase: se tienen en cuenta los 3 primeros trabajos realizados y la asistencia a clase. Como se deben buscar 3 valores posibles para este atributo, se tienen: Mucha participación: debe tener asistencia mayor o igual al 70% y el puntaje de los primeros 3 trabajos debe ser mayor o igual al 70%.

Media participación: debe tener asistencia mayor o igual al 70% y el puntaje de los primeros 3 trabajos debe ser menor al 70 % ó debe tener asistencia menor al 70% y el puntaje de los primeros 3 trabajos debe ser mayor o igual al 70 %.

Poca participación: debe tener asistencia menor al 70% y el puntaje de los primeros 3 trabajos menor al 70%.

## 6. Resultados

Los datos iniciales son utilizados como fuente para entrenar al algoritmo de decisión, es decir, el algoritmo se basa en esos registros previos cargados para poder analizar los nuevos datos ingresados, entre mayor volumen de datos sea tomado como entrenamiento, más preciso se vuelve el algoritmo de decisión puesto que cuenta con mayor cantidad de registros para tomarlos como ejemplo.

Una vez preparada la base de datos, se importan los datos de entrenamiento cargados previamente a un archivo de texto plano con extensión “.arff”, esta acción se realiza para que la librería *weka* que posee el algoritmo de decisión pueda leer la información contenida en el archivo, arme el árbol de decisión y se prepare para analizar los nuevos datos ingresados.

El archivo importado tiene como nombre la carrera del alumno, por ejemplo, los archivos con datos importados se llaman: informática.arff, contabilidad.arff, psicología.arff, derecho.arff o diseño.arff, para el caso de este ejemplo se llama informática.arff y su contenido es el siguiente:

```
@relation aprobar
@attribute participacion {medio, mucho, poco}
@attribute asistencia {mayor70, menor70}
@attribute trabaja {si,no}
@attribute aprueba {si, no}
@data
medio,mayor70,si,si
poco,menor70,si,si
poco,mayor70,si,si
poco,mayor70,no,no
```

La palabra clave `@relation` define la información deseada y los atributos por `@attribute`, seguidos del nombre del atributo y entre llaves son colocados los posibles valores que pueden recibir esos campos.

Por último, se utiliza el `@data` para incluir a todos los datos importados de la base de datos, delimitados por coma y ordenados uno al lado de otro hasta el final del registro, el salto de línea indica que se trata de un nuevo registro.

Una vez cargados los datos de entrenamiento, se ingresan los datos nuevos a modo de poder realizar la predicción de lo que se desea saber, para este trabajo lo que se quiere conocer es la probabilidad que posee el alumno de aprobar o no una determinada materia.

El algoritmo utilizado para este propósito es el árbol de decisión J48, que es el encargado de clasificar los datos y dar la solución para cada caso nuevo ingresado. Los nuevos registros procesados por el algoritmo, pasarán a formar parte de los datos de entrenamiento una vez que se vuelva a importarlos.

Por *web service* se envían los atributos de participación (Fig. 3), asistencia y trabajo en clase para que puedan ser procesados por el algoritmo de decisión utilizando el proceso “Calcular”.

**Figura 3.** Paso de parámetros al web service.

```
@WebMethod(operationName = "Calcular")
public String Calcular(@WebParam(name = "name") String name,
                      @WebParam(name = "participacion") String participacion,
                      @WebParam(name = "asistencia") String asistencia,
                      @WebParam(name = "trabaja") String trabaja) {
```

Los datos del archivo son leídos e instanciados:

**Figura 4.** Lectura del archivo *.arff*.

```
ConverterUtils.DataSource ds = new ConverterUtils.DataSource(name);
Instances ins = ds.getDataSet();
//System.out.println(ins.toString());
ins.setClassIndex(3);
```

Luego se crea el árbol J48 y es utilizada la función “.buildClassifier” para construir el esquema de clasificación basándose en los atributos instanciados. Se crea



una nueva instancia para definir el grupo de datos y setear los valores. Se define una colección la cual se completa con lo instanciado previamente (Fig. 5).

**Figura 5. Creación del árbol J48.**

```
J48 arbol = new J48();
arbol.buildClassifier(ins);

Instance nuevo = new Instance(3);
nuevo.setDataset(ins);
nuevo.setValue(0, participacion);
nuevo.setValue(1, asistencia);
nuevo.setValue(2, trabaja);

double probabilidades2[] = arbol.distributionForInstance(nuevo);
```

Finalmente se retorna el resultado que es almacenado en la variable “result” (Fig. 6) y guardado en la base de datos.

**Figura 6. Carga del resultado al alumno.**

```
String qry ="update atributos "
+ "set atri_resultado= '"+result+"'"
+ "where personas_pers_ci="+ci+"; "
```

A continuación (Fig. 7), se pueden apreciar los resultados obtenidos:

<b>Identificación</b>	<b>Nombre</b>	<b>Aprueba</b>	<b>Reprueba</b>
10.588.654	Aldo	68 %	32 %
10.635.813	David	70 %	30 %
10.315.896	Marcelo	82 %	18 %
10.197.684	Nancy	20 %	80 %
10.257.321	María	90 %	10 %
10.613.796	Alfredo	10 %	90 %
10.934.837	Deysi	32 %	68 %
10.715.628	Fabiola	64 %	32 %
10.358.643	Cinthia	55 %	45 %
10.678.912	Daniela	60 %	40 %
10.297.683	Enrique	65 %	35 %
10.762.842	Liz	64 %	36 %
10.345.475	Marion	50 %	50 %
10.762.323	Damián	31 %	69 %
10.489.731	Agustina	81 %	19 %
10.986.345	Felipe	73 %	27 %
10.267.126	Santiago	81 %	19 %
10.934.837	Renato	81 %	19 %

## **7. Conclusiones**

El resultado de esta investigación sirve como una herramienta de apoyo al docente y contribuye a fortalecer el proceso de enseñanza-aprendizaje, a través de la interacción con los alumnos.

Existen personas que son ajenas a la tecnología, entre ellos pueden estar los docentes, que en su día a día lidian con el avance tecnológico, como por ejemplo al conectarse a la página de la universidad, llamar lista de manera virtual, utilizar el aula virtual, entre otros. Este proyecto se abocó a ellos, para brindarles una herramienta de apoyo intuitiva y amigable.

Se ha logrado la integración de una serie de herramientas y componentes que conforman la solución planteada. Al comenzar esta investigación se propuso el objetivo de implementar una aplicación móvil de apoyo al docente que permita la evaluación previa al examen final de un estudiante utilizando árboles de decisión; al finalizar este proyecto se puede afirmar que este objetivo fue cumplido satisfactoriamente.

Las pruebas y los resultados logrados permiten llegar a la conclusión de que la solución sirve de apoyo al docente en el proceso de enseñanza-aprendizaje.

## 8. Referencias

- Barchino R., Gutiérrez J.M., García E., Hilera R. (2001). EDVI: Un sistema de apoyo a la enseñanza presencial basado en Internet. Adquirido de: [bioinfo.uib.es/~joemiro/aenui/procJenui/Demosweb/PDFs/baedv72.pdf](http://bioinfo.uib.es/~joemiro/aenui/procJenui/Demosweb/PDFs/baedv72.pdf).
- Cardona P. (2004). Aplicación de árboles de decisión en modelos de riesgo crediticio. Adquirido de: [http://www.emis.ams.org/journals/RCE/V27/V27\\_2\\_139Cardona.pdf](http://www.emis.ams.org/journals/RCE/V27/V27_2_139Cardona.pdf).
- Chesñevar C. (s.f.). Inducción de árboles de decisión. Adquirido de: [http://lidecc.cs.uns.edu.ar/~cic/datamining/downloads/04\\_dectr.pdf](http://lidecc.cs.uns.edu.ar/~cic/datamining/downloads/04_dectr.pdf).
- Cover, T. M., & Thomas, J. A. (1991). Entropy, relative entropy and mutual information. Adquirido de: <http://www.cs.columbia.edu/~vh/courses/LexicalSemantics/Association/Cover%26Thomas-Ch2.pdf>.
- Marquès, P. (2000). Los docentes: funciones, roles, competencias necesarias, formación. Adquirido de: [http://www.uaa.mx/direcciones/dgdp/defaa/descargas/docentes\\_funciones.Pdf](http://www.uaa.mx/direcciones/dgdp/defaa/descargas/docentes_funciones.Pdf).
- Noguera V. (2018). Aplicación para apoyo al docente en el proceso enseñanza-aprendizaje utilizando árboles de decisión. Trabajo de Conclusión de Carrera, Carrera de Ingeniería en Informática, Universidad Columbia del Paraguay. 2018. 42pp.
- Rokach L. & Maimon O. (2015). Data Mining with Decision Trees. Adquirido de: [https://doc.lagout.org/Others/Data%20Mining/Data%20Mining%20with%20Decision%20Trees\\_%20Theory%20and%20Applications%20%282nd%20ed.%29%20%5BRokach%20%26%20Maimon%202014-10-23%5D.pdf](https://doc.lagout.org/Others/Data%20Mining/Data%20Mining%20with%20Decision%20Trees_%20Theory%20and%20Applications%20%282nd%20ed.%29%20%5BRokach%20%26%20Maimon%202014-10-23%5D.pdf).
- Valero S., Vargas A., García M. (sin año). Minería de datos predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Adquirido de: <http://www.utim.edu.mx/~svalero/docs/e1.pdf>.