

**Módulo:** Inteligencia Artificial

**Autores**

- Ricardo Javier Esparza Torres
- Gerardo Alfonso Ramírez Mejía

# PASO A PASO PARA LA REPLICACIÓN DE LA INVESTIGACIÓN DE IA EN LA EDUCACIÓN

## 1. ARCHIVOS NECESARIOS

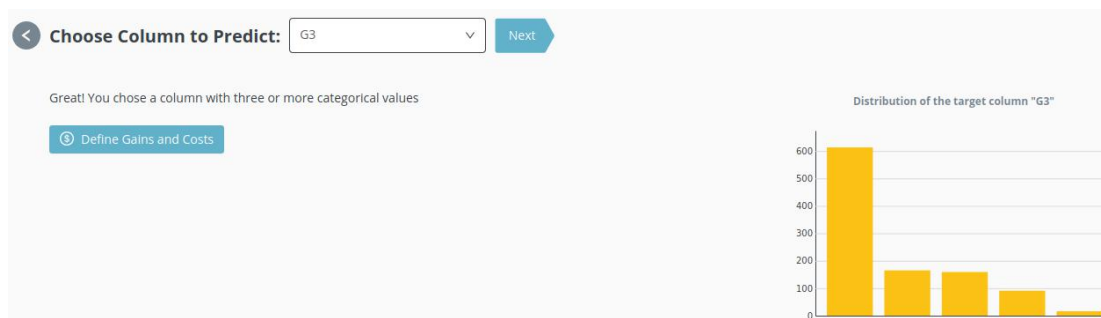
Para poder replicar la investigación se necesitan los archivos .csv originales o si lo desea el archivo .csv utilizado en la investigación. Los archivos originales se encuentran en el repositorio [UCI Machine Learning Repository](#) y el archivo editado lo pueden encontrar en nuestro repositorio de [GitHub](#).

## 2. PASO A PASO

Antes de dirigirnos a estas herramientas debemos tratar los dataset, es decir, analizarlos y limpiarlos de atributos o información innecesaria o implete, para esto se debe aplicar el proceso de análisis exploratorio.

### a) RAPIDMINER GO

Luego de analizar y limpiar los dataset los cargaremos dentro de rapidminer go, una vez cargado debemos elegir la columna que desamos predecir.



Ya seleccionada la columna a predecir, esta herramienta web nos indica que atributos son aquellos que tienen una calidad cuestionable o pobre y cuales tienen una correlación fuerte o debil.

En este paso podremos seleccionar o deseleccionar atributos.


Filter columns: <span>All (31)</span> <span>Questionable quality (2)</span> <span>Poor quality input (1)</span> <span>(30 of 31 columns selected)</span> <input type="text" value="Search Columns"/>						
	Name	Quality	Correlation	ID-ness	Stability	Missing
<input type="checkbox"/>	higher	Highly stable	0.92%	0.19%	91.48%	0.00%
<input checked="" type="checkbox"/>	G1	High correlation	46.91%	0.48%	59.58%	0.00%
<input checked="" type="checkbox"/>	G2	High correlation	57.84%	0.48%	57.47%	0.00%
<input checked="" type="checkbox"/>	school		0.20%	0.19%	73.95%	0.00%
<input checked="" type="checkbox"/>	sex		0.35%	0.19%	56.61%	0.00%
<input checked="" type="checkbox"/>	age		1.01%	0.77%	26.92%	0.00%
<input checked="" type="checkbox"/>	address		0.85%	0.19%	72.70%	0.00%
<input checked="" type="checkbox"/>	famsize		0.55%	0.19%	70.69%	0.00%
<input checked="" type="checkbox"/>	Pstatus		0.27%	0.19%	88.41%	0.00%

Continuamos al siguiente paso y podremos seleccionar los modelos que utilizares, en este paso es recomendable seleccionar “Higher Accuracy”, ya que esta opción comprueba varios modelos y nos muestra toda la información necesaria para poder tomar una decisión.

<
Select your models
▶ Run Analysis


Are you looking for quick insights? Or do you want easily interpretable models? Choose your model type

### Machine Learning Algorithms

☐



**Easily Interpretable**

Decision trees and simple linear models

☐


**Quick prototyping**

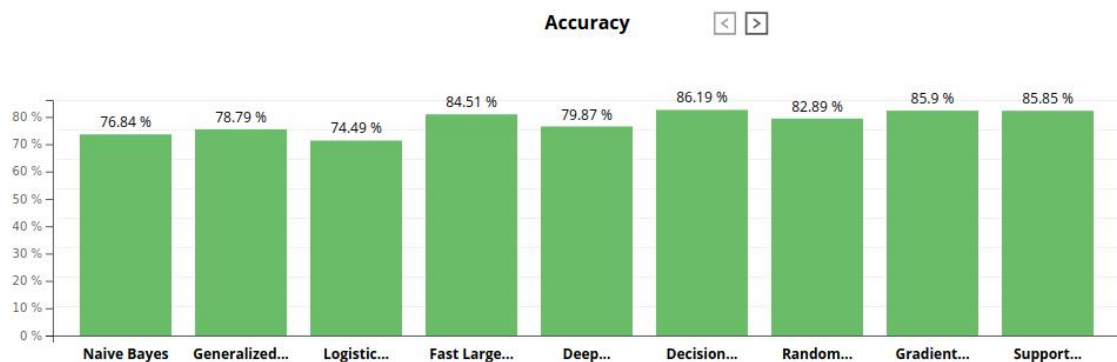
Quick insights using generic models

☒


**Higher Accuracy**

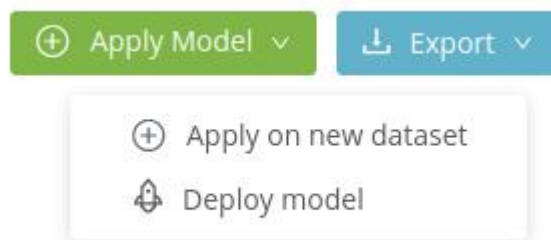
Longer running time with bigger selection of models

Esperamos un momento a que la herramienta ejecute varios modelos según nuestra data.



En la gráfica que nos presenta podemos ver la exactitud del modelo, esto es una gran ayuda para poder tomar la decisión de que modelos utilizar.

Seleccionamos el modelo a usar y rapidminer go en su versión de prueba nos permite deployar el modelo y utilizarlo.



Esto lo hace mediante un api y enviando la información necesaria en json.

Manage Deployment

X

This model is deployed. Click undeploy to revoke access to the model.  
You can call this URL using HTTP POST method.  
Don't paste the URL into the browser - use Postman or curl instead.

<https://go.rapidminer.com/am/api/deployments/f6626713-14aa-478f->

To score using the above model, send an HTTP POST request with your JSON encoded data in the body in the below format:

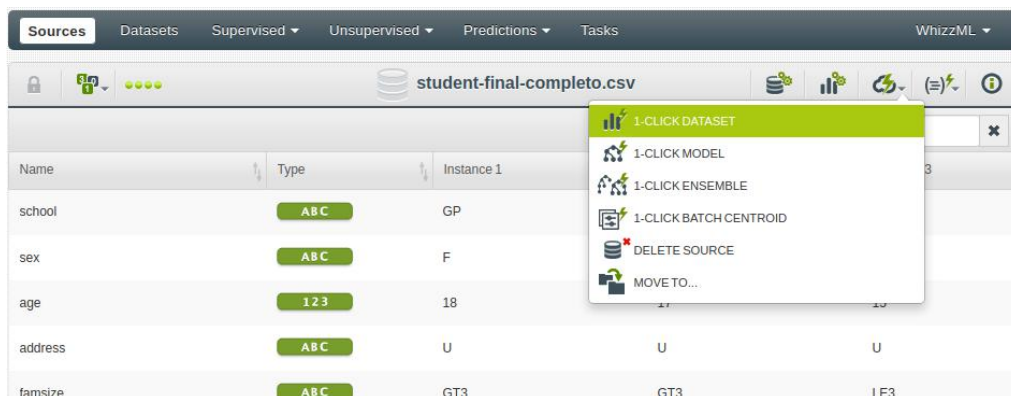
```
{
  "data": [
    {
      "school": "GP",
      "env": "E"
    }
  ]
}
```

Undeploy Model

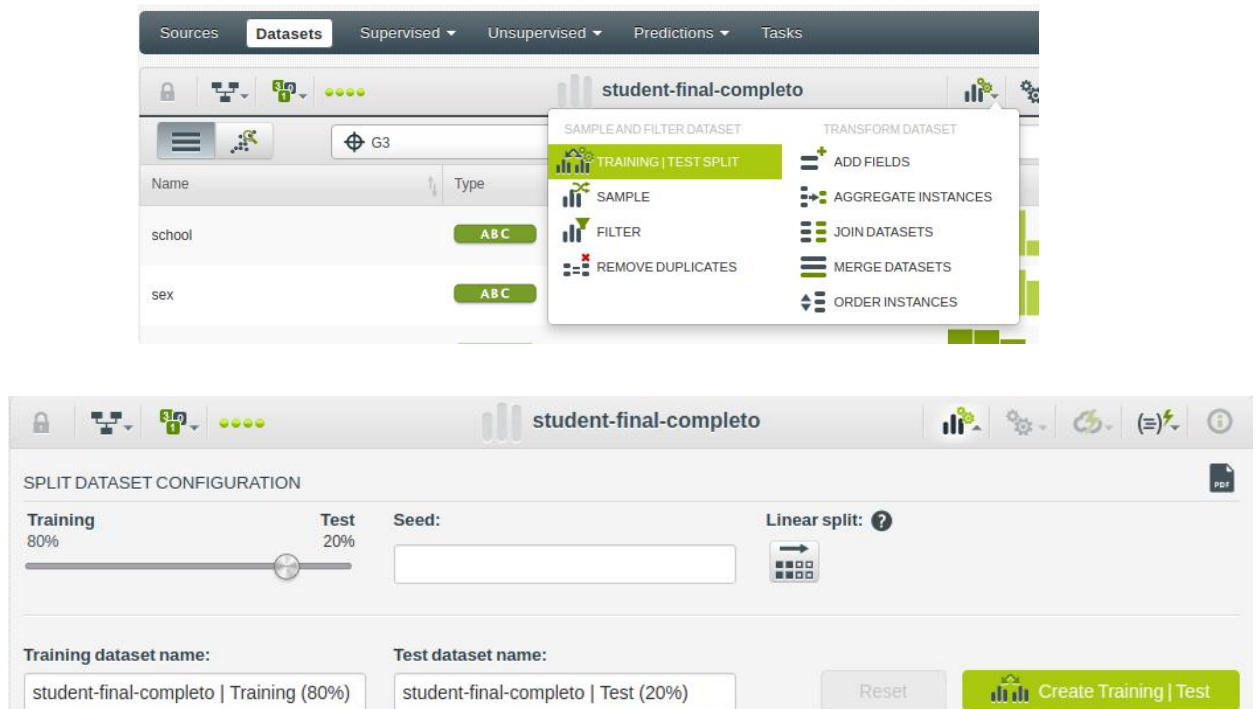
Close

## b) BIGML

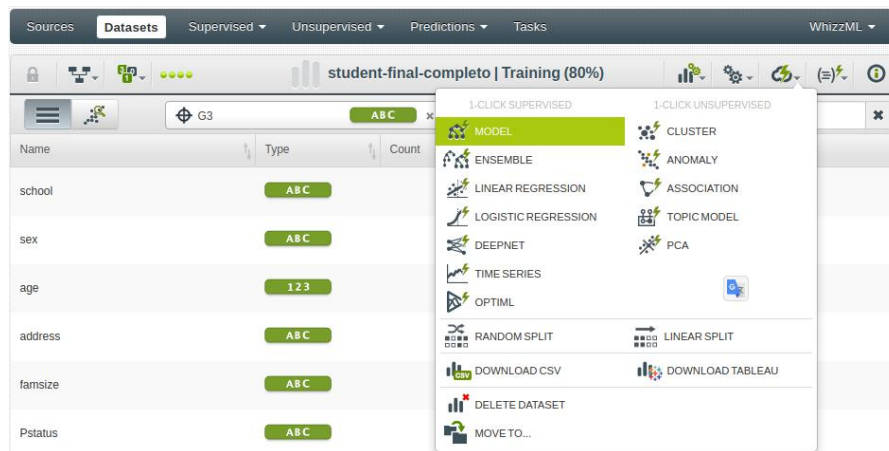
En BigML el proceso es un poco diferente, creada nuestra cuenta nos dirigimos a nuestro dashboard y ahí podremos cargar nuestro dataset. Una vez cargado, lo seleccionamos y como se muestra en la siguiente imagen convertimos nuestro archivo csv en un dataset.



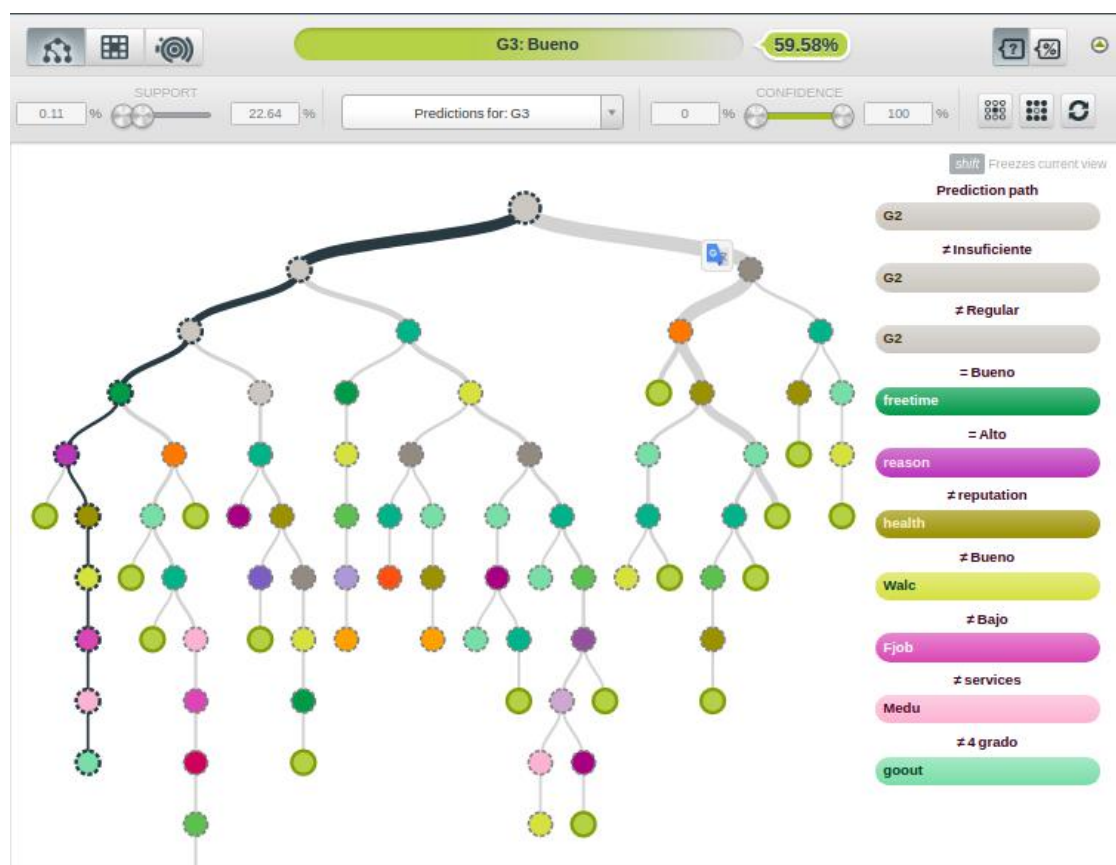
Después de convertir nuestra data en un dataset debemos dividirla en data de entrenamiento y data para las pruebas, lo recomendable es un 80% para entrenamiento y 20% para pruebas.



Como lo vimos en RapidMiner GO, el mejor modelo para nuestra data son los arboles de desición, entonces seleccionamos ese modelo en BigML.



Esperamos un momento a que la herramienta ejecute el modelo en nuestra data.



Ahora lo que debemos hacer es evaluar este modelo, lo cual haremos con nuestra data de pruebas.

student-final-completo | Training (80%)

176.6 KB size, 23 fields, 835 instances

student-final-completo | Test (20%)

44.2 KB size, 32 fields, 209 instances

Evaluation name: student-final-completo | Training (80%) vs. student-final-completo

Positive class: All classes

ACTUAL VS. PREDICTED	Buena	Insuficiente	Muy Buena	Regular	Sobresaliente	ACTUAL	RECALL	F	Phi
Buena	23	0	4	2	0	29	79.31%	0.73	0.69
Insuficiente	0	121	0	3	0	124	97.58%	0.96	0.90
Muy Buena	2	0	17	0	0	19	89.47%	0.85	0.84
Regular	9	7	0	19	0	35	54.29%	0.64	0.60
Sobresaliente	0	0	0	0	2	2	100.00%	1.00	1.00
PREDICTED	34	128	21	24	2	209	84.13% AVG. RECALL	0.84 AVG. F	0.80 AVG. Phi
PRECISION	67.65%	94.53%	80.95%	79.17%	100.00%	84.46% AVG. PRECISION	87.08% ACCURACY		

87.1% Accuracy

0.8369 F-measure

84.5% Precision

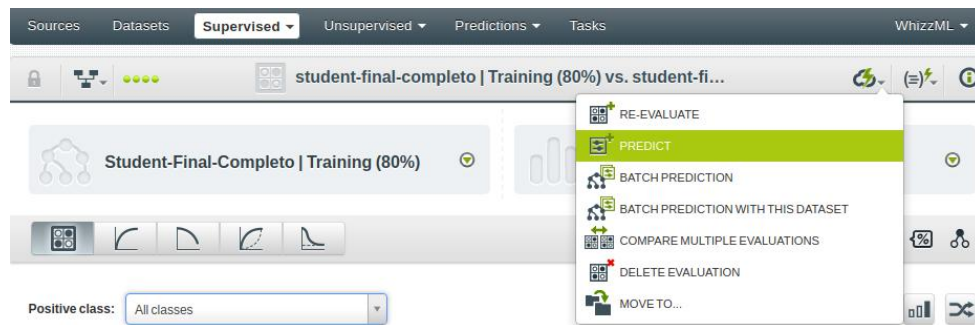
84.1% Recall

0.8048 Phi coefficient

Una vez evaluado podemos ver su nivel de exactitud que en este caso es del 87,1% y su presición es del 84,5%.



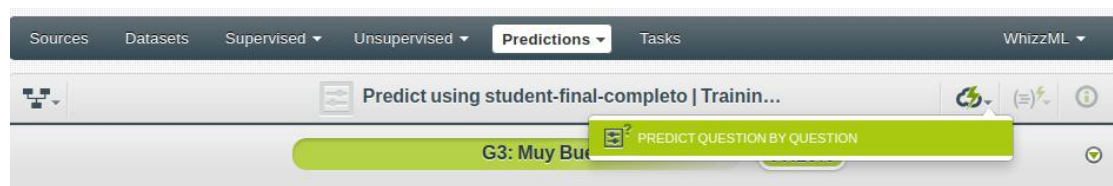
El siguiente paso es poder predecir el rendimiento académico, para esto BigML ofrece una herramienta para poder hacerlo.



Podemos predecir respondiendo a las preguntas necesarias como en un formulario.

A screenshot of the BigML prediction form. At the top, it says 'G3: Muy Bueno' with a green progress bar and '77.16%'. Below this, there are various input fields for different variables. On the left, under 'Missing strategy', there are icons for 'All input fields' and 'Select a positive class'. A 'Probability threshold' slider is set to 50%. The input fields are arranged in two columns. The left column includes: 'G2' (Buena, 91.80%), 'G1' (Buena, 6.27%), 'goout' (Alto, 4.71%), 'freetime' (Alto, 3.60%), and 'Dalc' (Alto, 2.03%). The right column includes: 'famrel' (Buena, 9.95%), 'Walc' (Alto, 5.37%), 'health' (Buena, 4.29%), 'age' (13 to 23, 16, 2.84%), and 'Medu' (4 grado, 1.55%). Each field has a green checkmark indicating it is completed.

Tambien BigML lo permite hacer pregunta a pregunta.



Predict using student-final-completo | Trainin...

G3: Muy Bueno

73.40%

Question:  
health?

Bueno

Start over

Next question

Answers:

G2? *Muy Bueno*

famrel? *Buena*

De esta manera tenemos dos formas para poder predecir el rendimiento de los estudiantes en su tercer trimestre segun la información recolectada.