

Manual de SDA-Vis: Sistema de visualización para el análisis de abandono de estudios

Nombre Apellido: GERMAIN GARCÍA ZANABRIA

Nombre Apellido 2

Nombre Apellido N

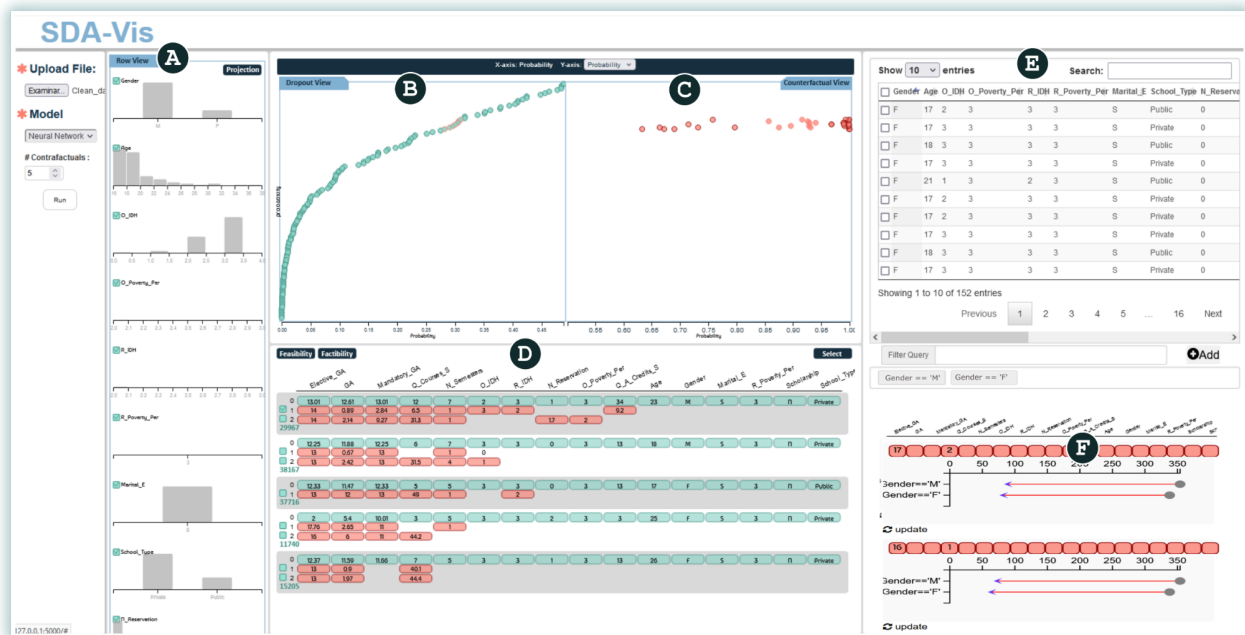
INDICE

Plataforma de análisis de datos de abandono estudiantil	2
Flujo de trabajo de exploración visual	3
Carga del conjunto de datos	4
Análisis basado en contrafactuales	5
Medir impacto de contrafactuales	8
Consulta de Imagen	9
Guardado de Imagen	12

Plataforma de análisis de datos de abandono estudiantil

El núcleo de SDA-Vis es la generación de escenarios sintéticos sin deserción utilizando contrafactuales con cambios menores en las características de los estudiantes.

Para ingresar a la plataforma se debe abrir el archivo de tipo **application** dentro de la carpeta del proyecto. Una vez abierto se mostrará la siguiente ventana:



La aplicación tiene los siguientes componentes visuales:

- A. Feature Distribution Bars view,
- B. Student Projection view,
- C. Counterfactual Projection view,
- D. Counterfactual Exploration view,
- E. Table view, y
- F. Impact view.

que hacen referencia a: (A) histogramas para representar la distribución de las características de los estudiantes, (B) una proyección para representar las probabilidades de los estudiantes según el umbral del modelo, (C) una proyección para representar los contrafactuales y sus probabilidades, (D) exploración contrafactual para representar los valores sintéticos, (E) una tabla que muestra todos los valores reales de los estudiantes que desertaron, y (F) una visualización del impacto de algunos contrafactuales en un determinado grupo de estudiantes. Los detalles de los componentes visuales se encuentran reportados en la memoria descriptiva.

Flujo de trabajo de exploración visual

El usuario primero carga el conjunto de datos, luego las características y distribuciones de los estudiantes se calculan y utilizan para guiar el análisis. Para este propósito, los usuarios emplean Feature Distribution Bars View **(A)**. Una vez que el usuario está familiarizado con los atributos, es necesario que el análisis de los potenciales desertores sea mapeado en la vista **SP B**. El usuario puede seleccionar un conjunto diferente de estudiantes en función a los indicadores (por ejemplo, probabilidad, viabilidad y factibilidad). Una vez identificado un grupo de interés en la vista **SP**, nuestra vista **CP (C)** muestra todos los contrafactuales asociados con la anterior selección. El usuario puede elegir libremente un conjunto de contrafactuales para inspeccionar. Para dicha inspección, el siguiente paso es utilizar Counterfactual Exploration view **(D)**, que muestra los valores originales para cada atributo. Al mismo tiempo, los contrafactuales calculados muestran sólo los valores que necesitan ser modificados. Los estudiantes y sus contrafactuales podrían ser reordenados en función de la viabilidad y factibilidad, para ayudar a los usuarios a seleccionar contrafactuales específicos para un mayor análisis e investigación. Finalmente, el usuario podría medir cómo los cambios sugeridos por un contrafactual seleccionado puede influir en algunos estudiantes que usan el Impacto (Impact view **(F)**). Esta visualización muestra hasta qué punto sería posible reducir el número de abandonos. Para ello, es necesario seleccionar subconjuntos de alumnos para los que el usuario quiera medir el impacto. Para elegir los subconjuntos, el usuario puede utilizar el Table View E, aplicando algunos filtros.

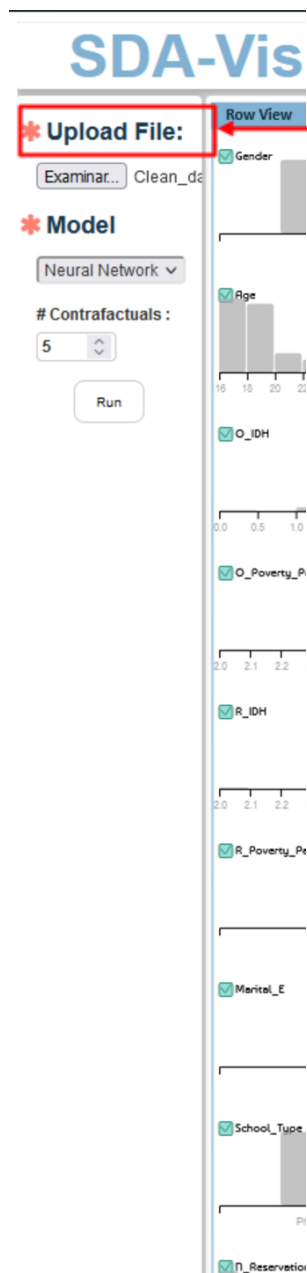
Si el análisis resultante no satisface las necesidades de los usuarios, los usuarios pueden seleccionar otros contrafactuales desde Counterfactual Exploration view o comenzar de nuevo desde Dropout Analysis dual view (**B** y **C**) para considerar un nuevo grupo de estudiantes/contrafactuales y obtener un análisis más detallado

A continuación algunas secciones en la plataforma serán descritas más detalladamente.

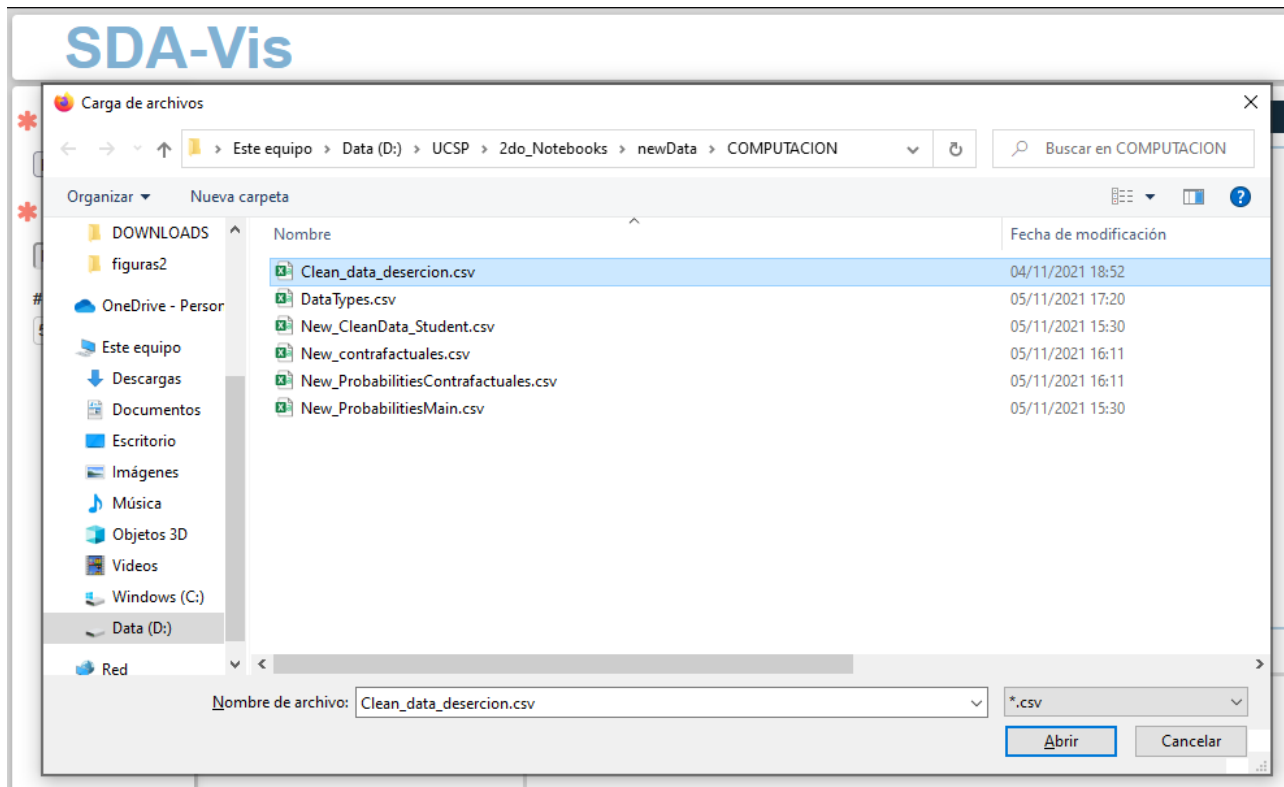
Carga del conjunto de datos

La función de esta sección en la plataforma es de seleccionar un conjunto de datos en alguna ubicación en el ordenador para poder procesar en la plataforma. El sistema verificará que la fuente de datos cumpla con los requisitos necesarios para ser válida, de lo contrario el archivo de datos no será aceptado y un mensaje de advertencia será mostrado.

El primer paso es hacer click en el botón “Cargar Imagen”, en la parte superior izquierda de la ventana.



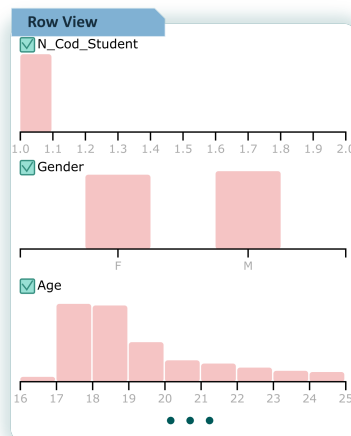
Luego, procedemos a seleccionar el conjunto de datos.



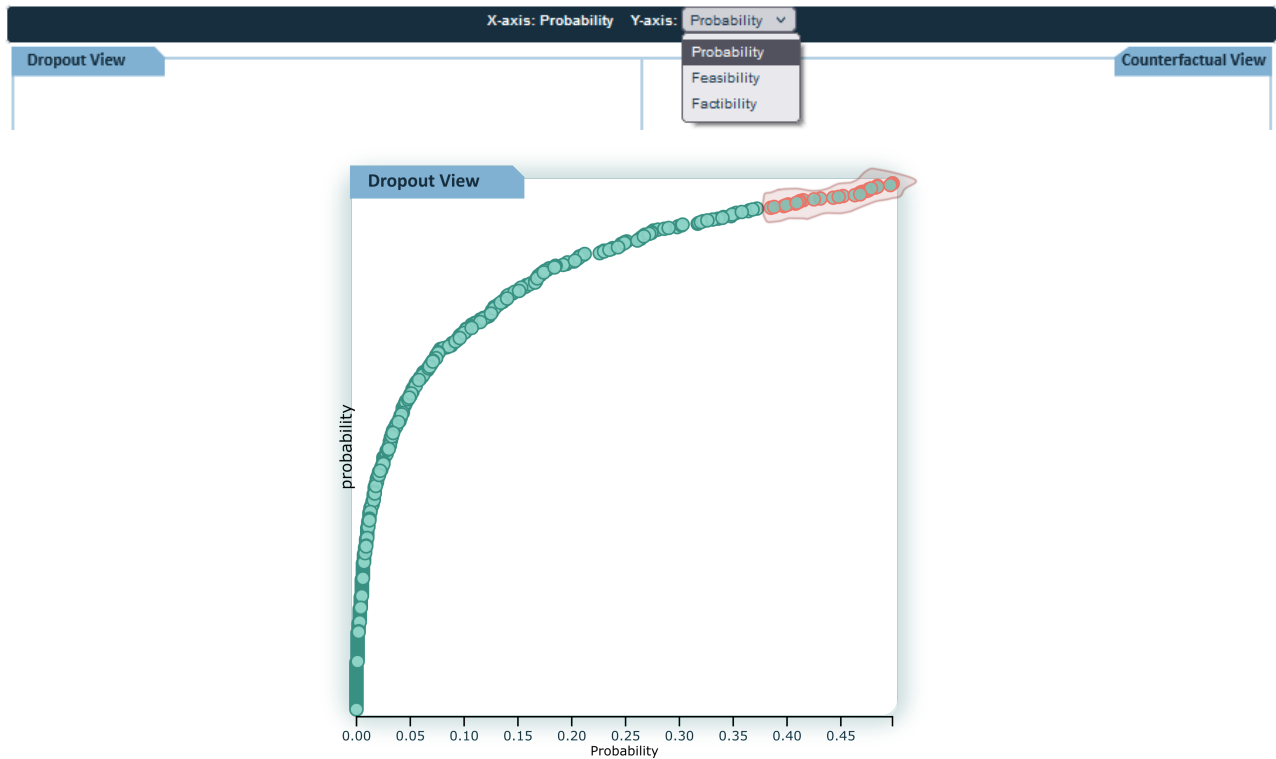
Una vez seleccionada la imagen, de ser válida el conjunto de datos se mostrará las visualizaciones posteriores con datos. Veamos esto con un ejemplo práctico.

Análisis basado en contrafactuales

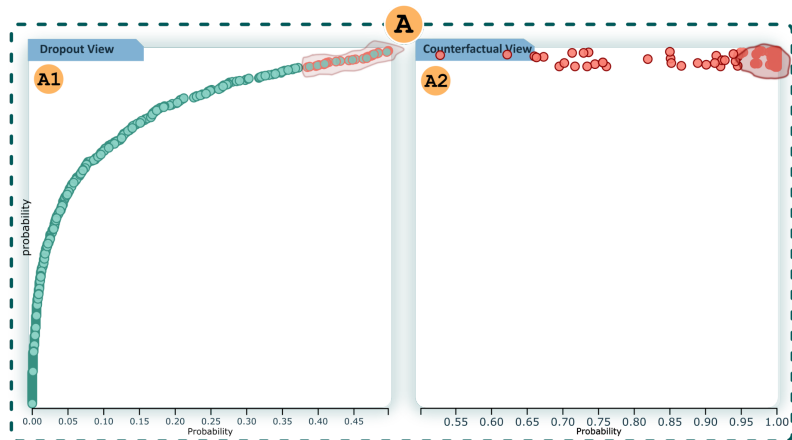
La primera visualización muestra la distribución de lo valores a estudiar:



La segunda visualización es la distribución de los estudiantes basado en su probabilidad:



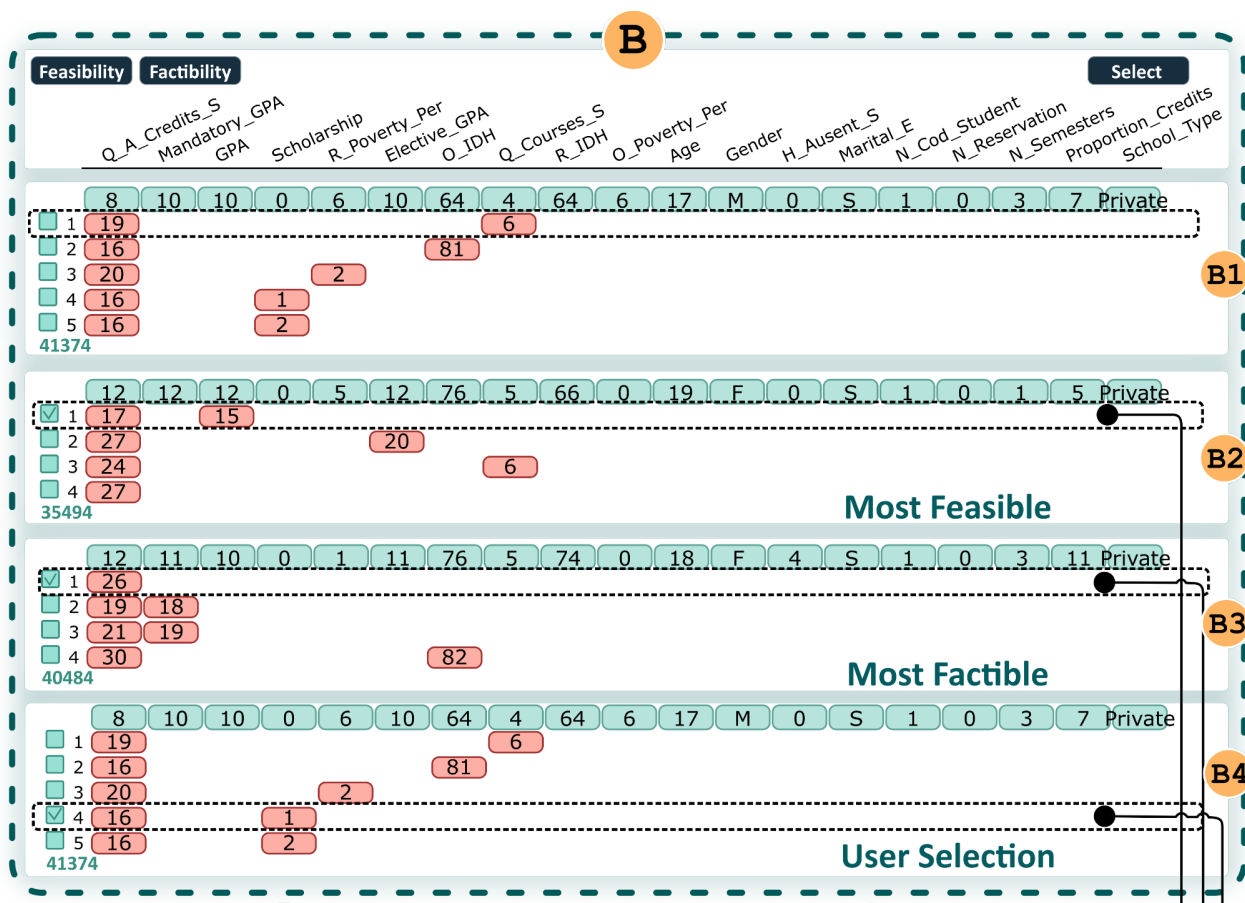
En esta visualización el usuario puede seleccionar bajo qué criterios quiere representar los datos: Probability, Feasibility y Factibility. Como siguiente paso se eligen los alumnos. La segunda visualización es la distribución de los estudiantes basado en su probabilidad para mostrar los contrafactuales.



La figura (A) muestra el *Dropout Analysis Dual-View* de los estudiantes y contrafactuales involucrados en el análisis. Basándose en la presentación por defecto de estas metáforas, el usuario puede redefinir su análisis, en este caso, se toma en cuenta los estudiantes cerca de la línea umbral (alta probabilidad) y qué variables deberían mejorarse para evitar la deserción. Usando la selección de lazo, se seleccionó a los estudiantes con la mayor probabilidad de no abandonar la universidad (resaltados con una polilínea en A1) esperando que necesiten menos cambios en sus valores originales de no abandono. Esta selección consta de 30 estudiantes,

Por otro lado, la Figura **A2** muestra alrededor de 150 contrafactuales (cinco por cada estudiante). Los contrafactuales también están ordenados por sus probabilidades; por lo tanto, su análisis se puede centrar en ciertos grupos en función de sus probabilidades. Analizando el contrafactual, *Counterfactual exploration view (A1)* evidenciado que las variables más relevantes están vinculadas a algunas variables (*Q_A_Credits_S*, *Mandatory_GPA*, *GPA*, *Scholarship*, *R_Poverty_Per*, *pobreza*, *Elective_GPA*, *O_HDI*, *Q_Courses_S*, etc). Por ejemplo, la primera alternativa para el primer estudiante (con guiones en **B1**) es aumentar el número de créditos aprobados por semestre (de 8 a 19) y la cantidad de cursos en un semestre (de 4 a 6). En la misma línea, es posible analizar todo los contrafactuales.

Sin embargo, analizar los contrafactuales individuales de cada estudiante es tedioso, lo que hace que, utilizando algunas organizaciones basadas en algunas métricas, sea una mejor alternativa.



Con un simple clic en el botón **Factible**, es posible reordenar a los estudiantes y su contrafactual en función de su viabilidad. La Figura **B2** muestra al estudiante con la alternativa más factible sugiriendo la posibilidad en *Q_A_Credits_S* de 12 a 17 y *GPA* de 12 a 15. En la misma línea, es posible recurrir a contrafactuales basados en su *Factibility*, como se muestra en **B3**.

Medir impacto de contrafactuales

Como siguiente paso podrían afectar a grupos específicos de estudiantes. Para este ejemplo simple, vamos a considerar grupos de género, seleccionamos grupos masculinos y femeninos, para eso, se usa Table view usando consultas de pandas ($Gender == 'M'$ y $Gender == 'F'$).

Show **10** entries Search:

<input type="checkbox"/>	Gender	Age	O_IDH	O_Poverty_Pér	R_IDH	R_Poverty_Pér	Marital_E	School_Type	N_Reserve
<input type="checkbox"/>	F	17	2	3	3	3	S	Public	0
<input type="checkbox"/>	F	17	3	3	3	3	S	Private	0
<input type="checkbox"/>	F	18	3	3	3	3	S	Public	0
<input type="checkbox"/>	F	17	3	3	3	3	S	Private	0
<input type="checkbox"/>	F	21	1	3	2	3	S	Public	0
<input type="checkbox"/>	F	17	2	3	3	3	S	Private	0
<input type="checkbox"/>	F	17	2	3	3	3	S	Private	0
<input type="checkbox"/>	F	17	3	3	3	3	S	Private	0
<input type="checkbox"/>	F	18	3	3	3	3	S	Public	0
<input type="checkbox"/>	F	17	3	3	3	3	S	Private	0

Showing 1 to 10 of 152 entries

Previous **1** 2 3 4 5 ... 16 Next

Filter Query + Add

Gender == 'M' Gender == 'F'

Como paso final usamos Impact view para mostrar el impacto de cada contrafactual seleccionado. En este caso, el contrafactual más factible (D1) y sus cambios sugeridos logran reducir los 77.4% de deserción masculina, 79.3% de deserción femenina y 78.3% de abandono total. De la misma forma, el contrafactual más factible es reducir los abandonos masculinos en un 32.2%, los abandonos femeninos en un 26% y un 29% del total de abandonos. Finalmente, el contrafactual seleccionado de el asistente da el mejor resultado, reduciendo un 79.9% para los hombres, 80.7% para las mujeres y 79.9% del total de abandonos.

