

Instalación de Altair AI Studio®

Altair AI Studio® brinda una licencia de uso gratuito con la limitación de uso para no más de 10,000 registros; se puede descargar de:

<https://my.rapidminer.com/nexus/account/index.html#downloads>

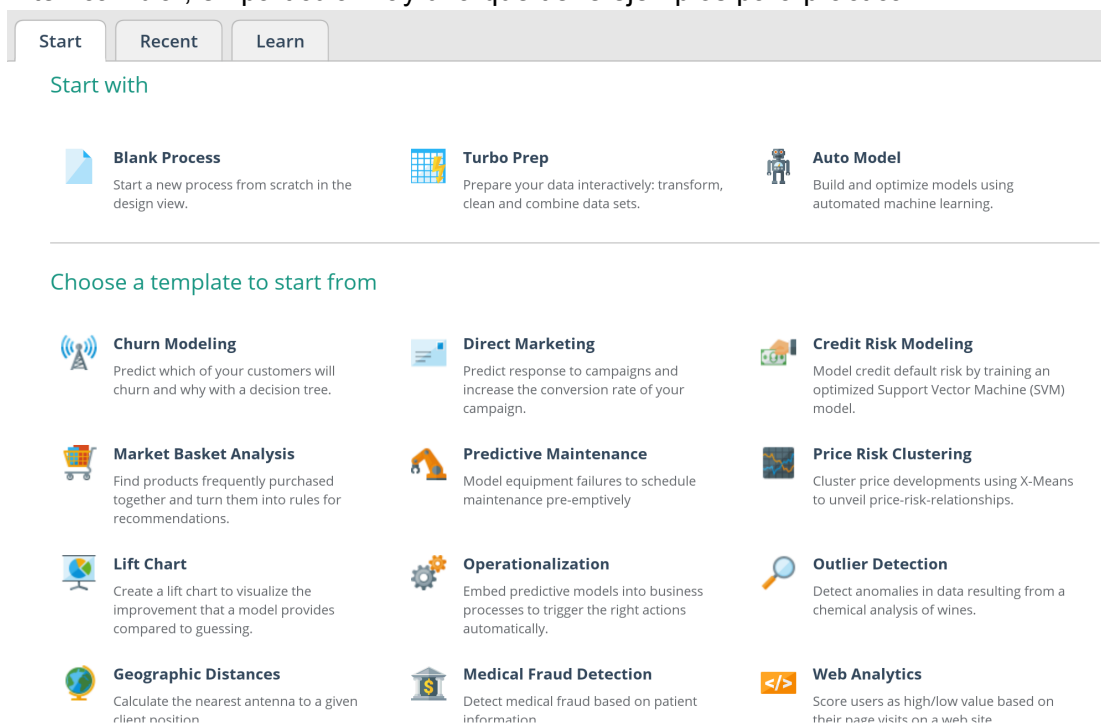
Para generar y activar la licencia, ingresar a:

<https://community.altair.com/discussion/33624/steps-to-install-and-activate-altair-ai-studio-for-educational-use-from-rapidminer-com-with-rapidminer-licensing-personal-edition>

Al crear la cuenta asegúrese de elegir Educational purposes (e.g., educator, student) para la licencia de uso gratuito.

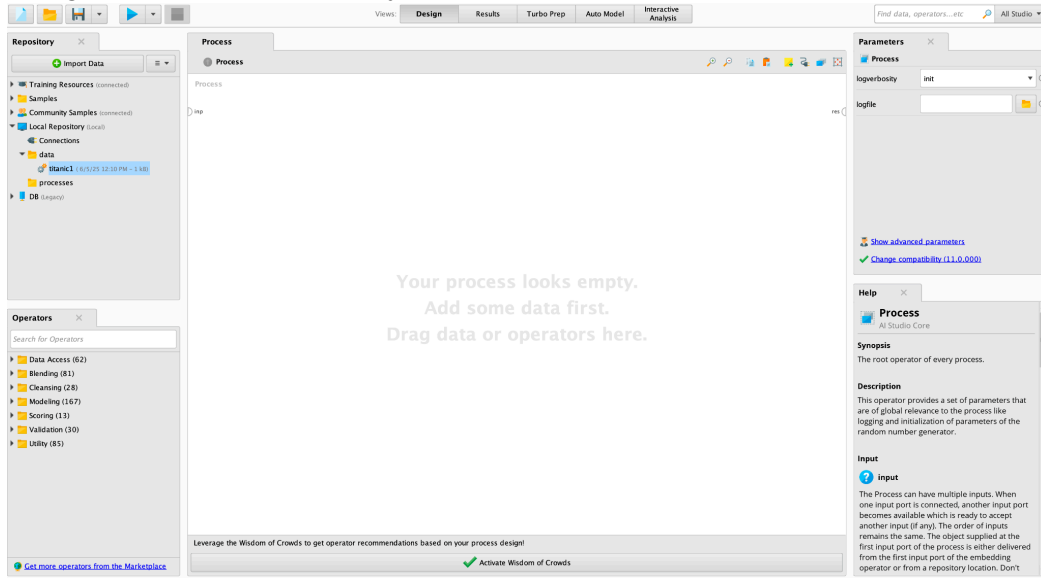
Ejemplo de uso de Altair AI Studio®

Al iniciar el programa nos presenta opciones para trabajar; las pestañas superiores nos permiten cambiar, en particular hay una que tiene ejemplos para practicar.

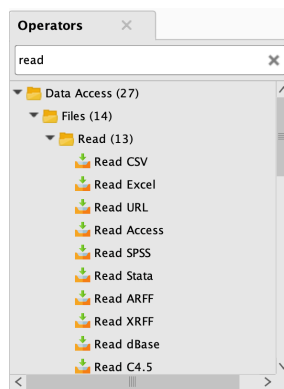


Este *software* permite realizar procesamiento completo del tipo *pipelines* de forma gráfica. Veamos un ejemplo sencillo de su uso.

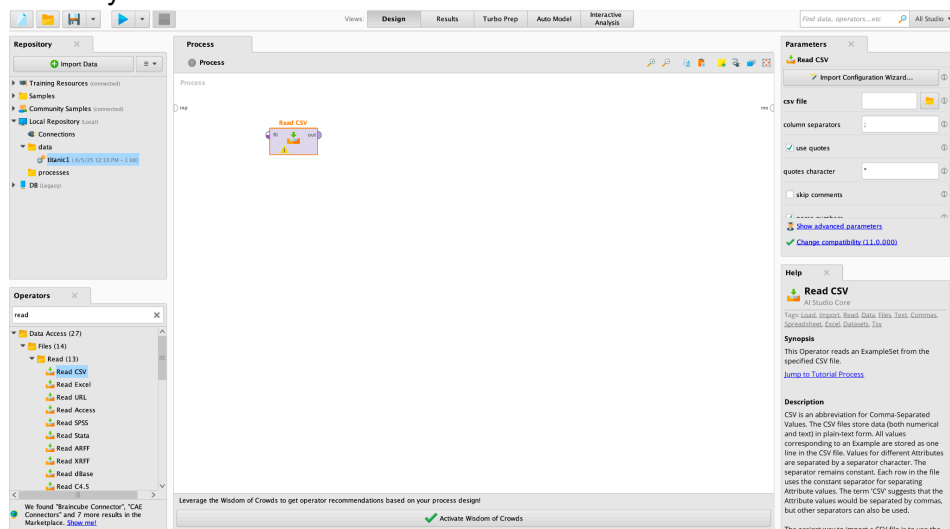
1. Elegimos un *Blank Process* y nos manda al la pantalla de inicio:



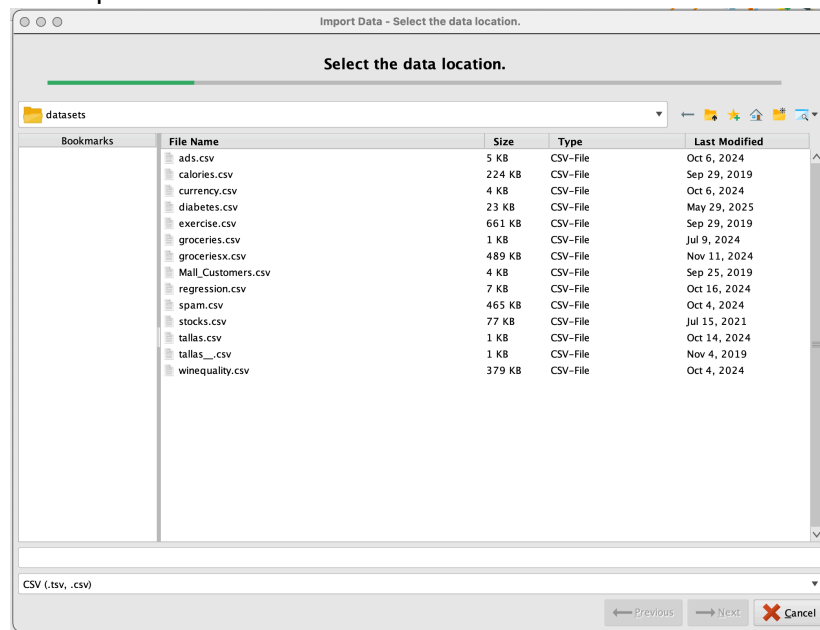
2. En la sección Operators podemos buscar “read” y nos presenta todas las opciones de lectura de información:



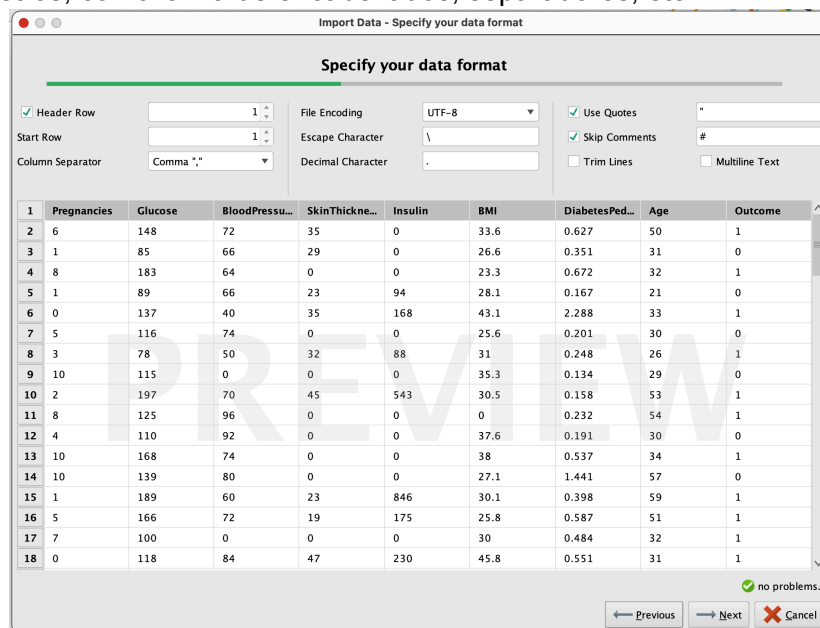
3. Elegimos *Read CSV* y lo arrastramos al área de proceso, es importante notar los cambios en el ambiente, por ejemplo en *Parameters* aparece que nos ayudará a :



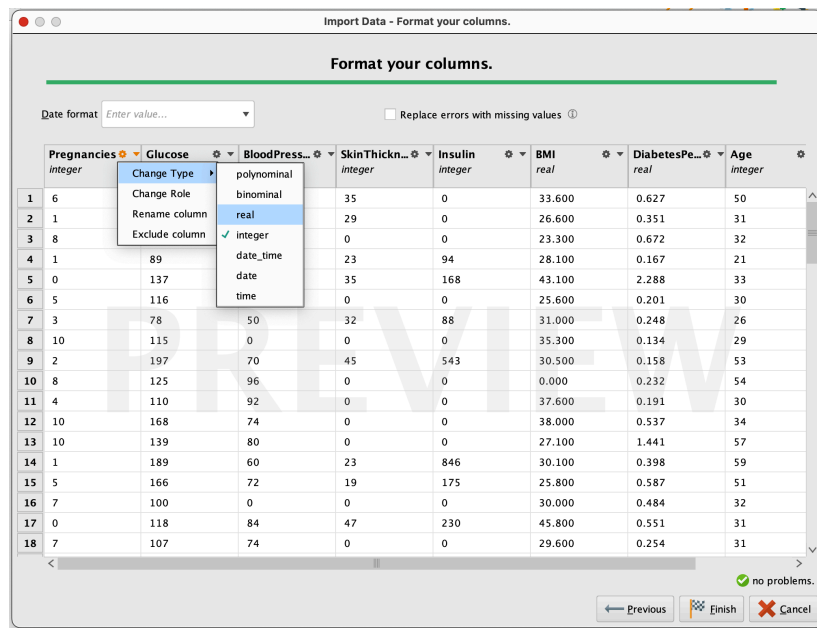
4. Al dar clic en el botón se abre una pantalla que permite seleccionar el archivo con los datos a importar:



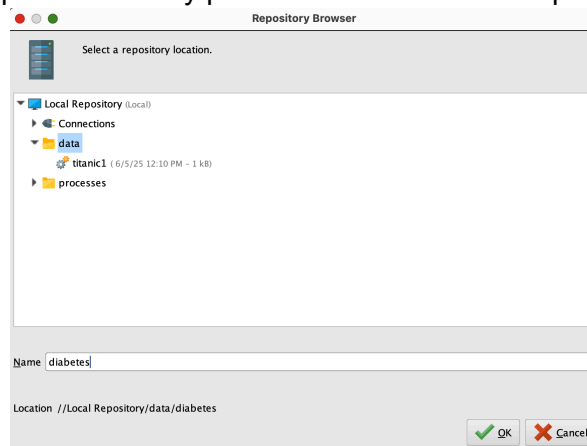
5. Elegimos el archivo, en este caso *diabetes.csv* y clic en *Next*, nos presenta una pantalla con una vista previa de nuestros datos, se pueden configurar algunos parámetros, como la fila de encabezados, separadores, etc.:



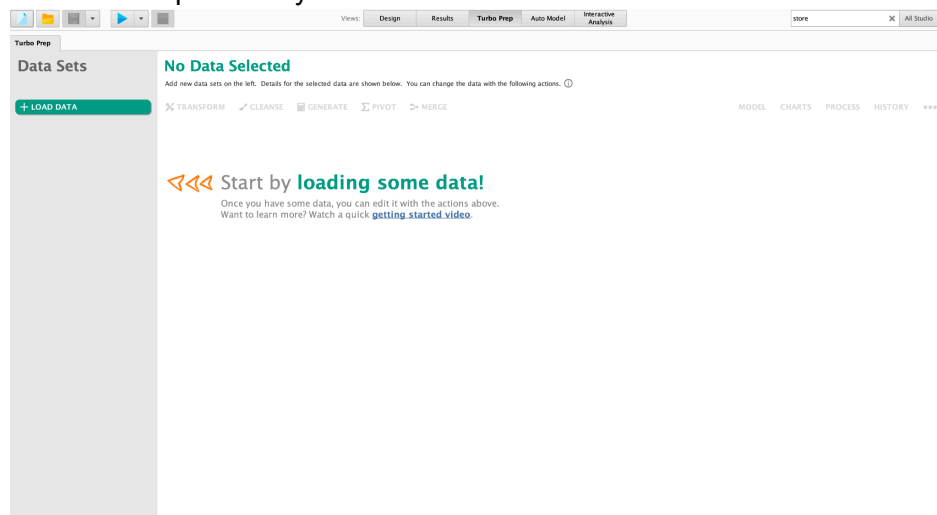
6. Al dar clic en *Next*, nos permite elegir y modificar algunas características de las columnas de nuestro conjunto de datos:



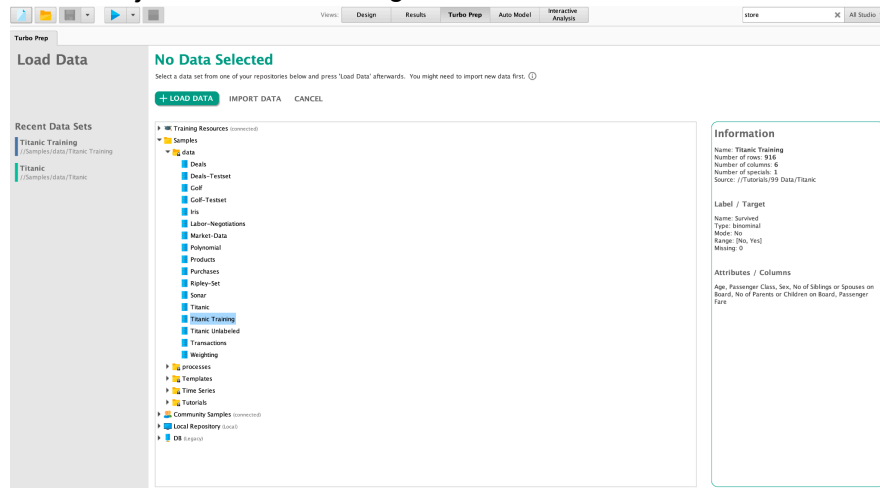
7. Damos clic en *Finish* y el conjunto de datos ya está disponible para nuestro uso. Podemos guardar nuestro proceso en *Local Repository* → *data*; es recomendable separar los bloques de datos y procesos en diferentes carpetas:



Ahora podemos pasar a realizar procesos con los datos; con el botón *Turbo Prep*, nos movemos a una zona que nos ayuda con diferentes acciones sobre los datos:



1. Damos clic en + **LOAD DATA** a la izquierda y podemos elegir de los datos de muestra el conjunto *Titanic Training* dando doble clic:



2. Nos muestra los datos con varias opciones como **TRANSFORM**, **CLEANSE**, etc.:

Age	Passenger Class	Sex	No of Siblings or Spouses on Board	No of Parents or Children on Board	Passenger Fare	Survived
29	First	Female	0	0	211.338	Yes
2	First	Female	1	2	151.550	No
30	First	Male	1	2	151.550	No
25	First	Female	1	2	151.550	No
48	First	Male	0	0	26.550	Yes
63	First	Female	1	0	77.958	Yes
39	First	Male	0	0	0	No
18	First	Female	1	0	227.525	Yes
26	First	Female	0	0	78.850	Yes
80	First	Male	0	0	30	Yes
29.881	First	Male	0	0	25.925	No
24	First	Male	0	1	247.521	No
50	First	Female	0	1	247.521	Yes
32	First	Female	0	0	76.292	Yes
36	First	Male	0	0	75.242	No
37	First	Male	1	1	52.554	Yes
26	First	Male	0	0	30	Yes
42	First	Female	0	0	227.525	Yes
29	First	Female	0	0	221.779	Yes
25	First	Male	1	0	91.079	Yes

3. Por ejemplo, podemos agregar columnas con **GENERATE**, por ejemplo el número de familiares de cada muestra sumando dos columnas que se pueden arrastrar directamente:

Formula: `No of Siblings or Spouses on Board + No of Parents or Children on Board`

UPDATE PREVIEW - Update the preview to make sure that all calculations can be performed.

No of Relatives	No of Siblings	No of Parents	Age	Passenger Class	Sex	Passenger Fare
0	0	0	29	First	Female	211.338
3	1	2	2	First	Female	151.550
3	1	2	30	First	Male	151.550
3	1	2	25	First	Female	151.550
0	0	0	48	First	Male	26.550
1	1	0	63	First	Female	77.958
0	0	0	39	First	Male	0
1	1	0	18	First	Female	227.525
0	0	0	26	First	Female	78.850

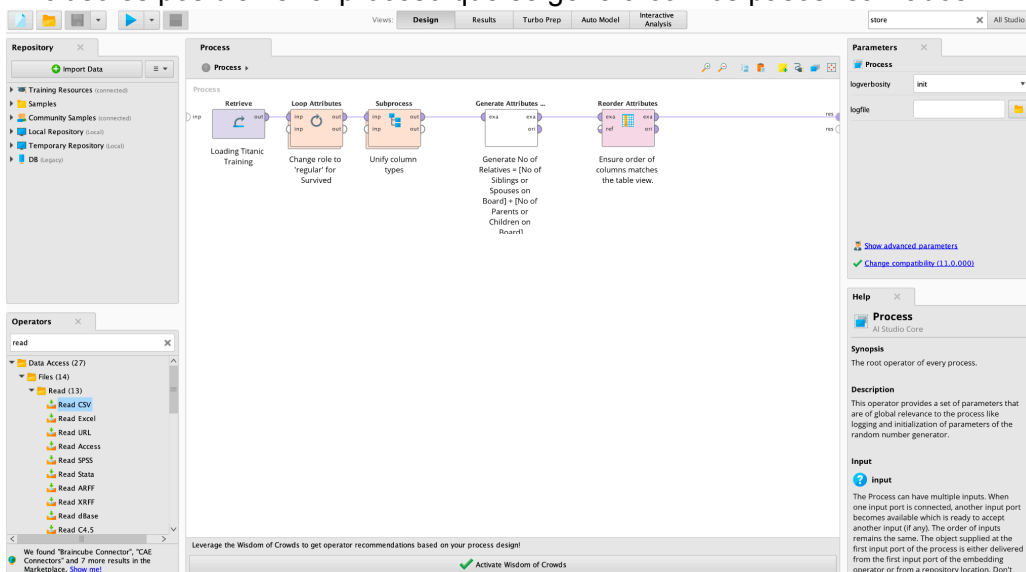
4. Damos clic en **UPDATE PREVIEW** y después en **COMMIT GENERATE**, tenemos nuestra nueva columna:

No of Relatives Number	No of Siblings or Spouses on Board	No of Parents or Children on Board	Age Number	Passenger Class Category	Sex Category	Passenger Fare Number	Survived Category
0	0	0	29	First	Female	211.338	Yes
3	1	2	2	First	Female	151.550	No
3	1	2	30	First	Male	151.550	No
3	1	2	25	First	Female	151.550	No
0	0	0	48	First	Male	26.550	Yes
1	1	0	63	First	Female	77.958	Yes
0	0	0	39	First	Male	0	No
1	1	0	18	First	Female	227.525	Yes
0	0	0	26	First	Female	78.850	Yes
0	0	0	80	First	Male	30	Yes
0	0	0	29.881	First	Male	25.925	No
1	0	1	24	First	Male	247.521	No
1	0	1	50	First	Female	247.521	Yes
0	0	0	32	First	Female	76.292	Yes
0	0	0	36	First	Male	75.242	No
2	1	1	37	First	Male	52.554	Yes
0	0	0	26	First	Male	30	Yes
0	0	0	42	First	Female	227.525	Yes
0	0	0	29	First	Female	221.779	Yes
1	1	0	76	First	Male	61.078	Yes

5. Podemos ver el historial de cambios realizados con **HISTORY**, ahí se puede elegir deshacer alguna transformación realizada



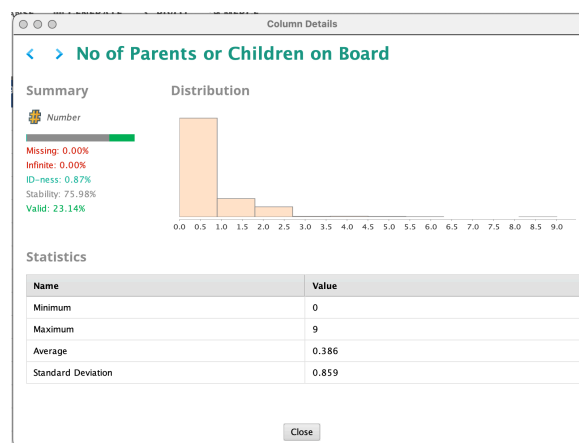
6. Incluso es posible ver el proceso que se genera con los pasos realizados:



7. En la vista de los datos, se puede dar clic sobre alguna columna y se presentan también opciones de transformación y limpieza, por ejemplo ordenar los datos:

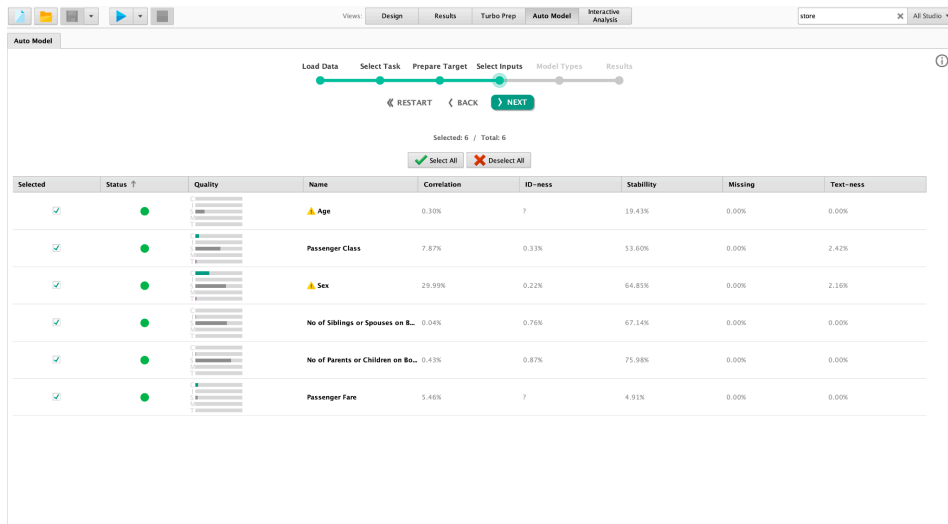
The screenshot shows the 'Titanic Training' dataset in the Data Studio interface. The 'No of Parents or Children on Board' column is selected, and a context menu is open. The 'Sort View (Descending)' option is highlighted, and a tooltip indicates 'Applies the transformation Sort'.

8. Con *Show Details* obtenemos una vista más profunda sobre la columna seleccionada:

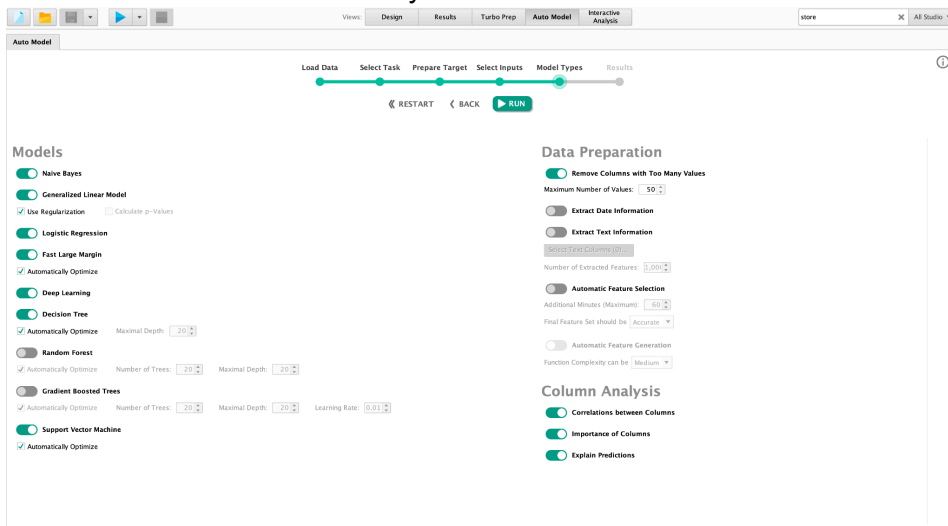


9. Así mismo, es posible crear un modelo rápido, dando clic en *CREATE PROCESS* o directamente en el botón *AUTO MODEL* se selecciona el tipo de modelo *Predict* con la columna *Survived* y como objetivo, luego *NEXT* seleccionando todas las columnas:

The screenshot shows the 'Auto Model' interface. The 'Predict' task is selected, and the 'Survived' column is chosen as the target. The 'NEXT' button is highlighted, indicating the next step in the model creation process.



10. Para este ejemplo, omitimos *Random Forest* y *Gradient Boosted Trees* para que la ejecución no demore demasiado y damos clic en **RUN**:



11. Y nos regresa un resumen con los resultados de cada modelo, así como la indicación del mejor modelo generado, así como métricas como la curva ROC:

