****

**Laboratorio :** AutoML con Watson Studio

María Borbonés García

@mariaborbones

Software,acceso y archivos necesarios 3

Parte 1: Generación de pipelines AutoML 4

AutoAI en SPSS Modeler 27

Modelado automático 27

Selección de características 37

Preparación automática de datos 39

En este laboratorio aprenderá a usar AutoAI en Watson **Studio**. AutoAI puede mejorar la productividad y la calidad de los modelos de los científicos de datos. Para obtener más información acerca de AutoAI, consulte la página del producto: <https://www.ibm.com/cloud/watson-studio/autoai>

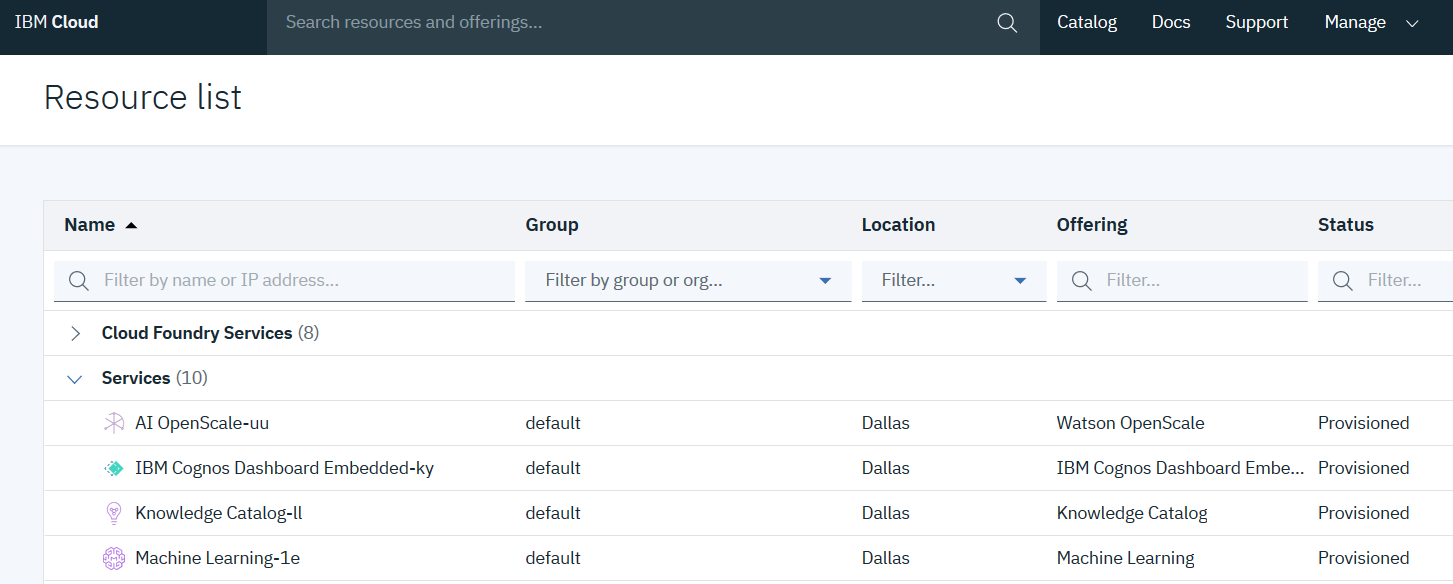
# Software,acceso y archivos necesarios

Para completar este laboratorio, necesitarás una cuenta de Watson Studio Cloud**:** <https://dataplatform.cloud.ibm.com>

* Si no tiene una cuenta de Watson Studio, use la misma dirección URL para registrarse para obtener una prueba gratuita. La cuenta se activará en aproximadamente 5 minutos.

Además de Watson Studio, necesitará Watson **Machine Learning** **Service** (**WML**).

* + Si acaba de crear una cuenta de prueba, el servicio WML se aprovisionó automáticamente.
  + Si tiene que aprovisionarlo manualmente, puede hacerlo desde su **Dashboard** de IBM Cloud: <https://cloud.ibm.com/login>
    - Compruebe si el servicio WML se muestra en **Servicios**. Si no es así, búsquelo en el **catálogo** y agréguelo.

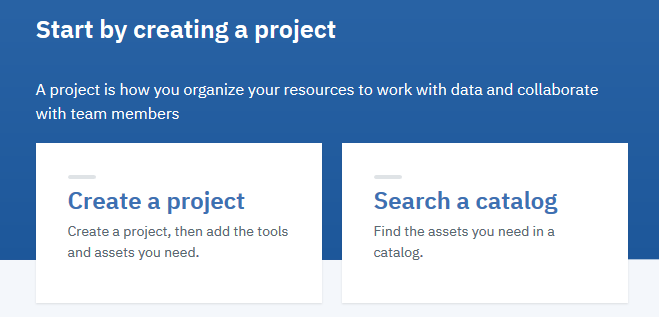


También necesitarás archivos de esta página GitHub: <https://github.com/mariaborbones/ejercicios_IA>

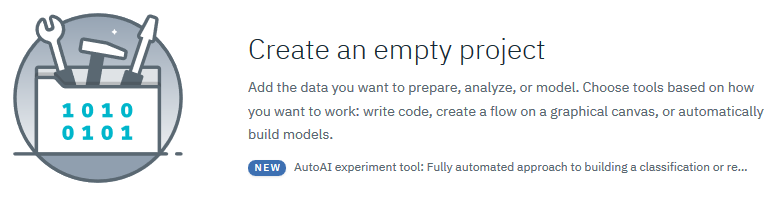
# Parte 1: Generación de pipelines AutoML

**Watson Studio** proporciona varias capacidadesde *AutoAI.* En esta sección aprenderá a usar Watson **Studio** para crear automáticamente modelos de clasificación.

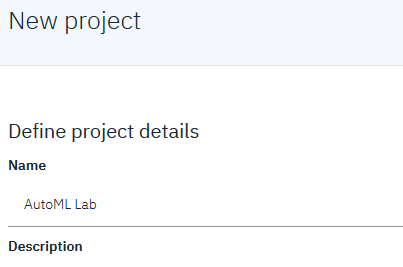
1. Inicie sesión en **Watson Studio Cloud**.
2. En la pantalla de bienvenida, haga clic en **Crear un proyecto**



1. Seleccione **Crear un proyecto vacío**.

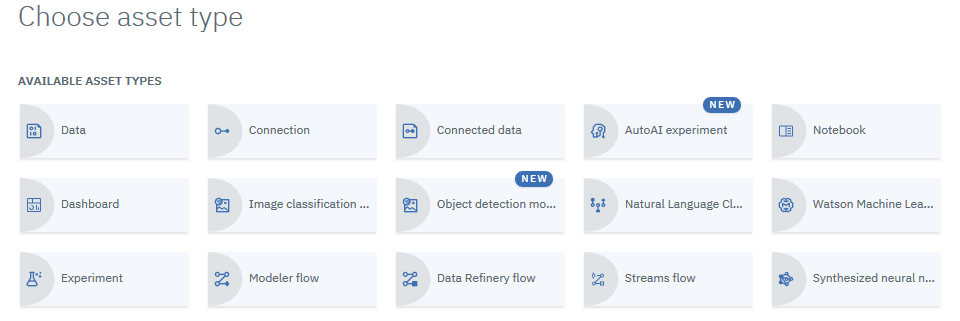


1. Asigne al proyecto el nombre **AutoML Lab** y haga clic en **Crear**.

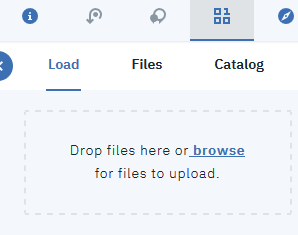


1. A continuación, cargaremos los activos de datos que se usarán para el modelado. Los archivos .csv se encuentran en el directorio **/data** del archivo que descargó de **GitHub**.
   1. En **Project** view seleccione Agregar al proyecto **-> Datos**.

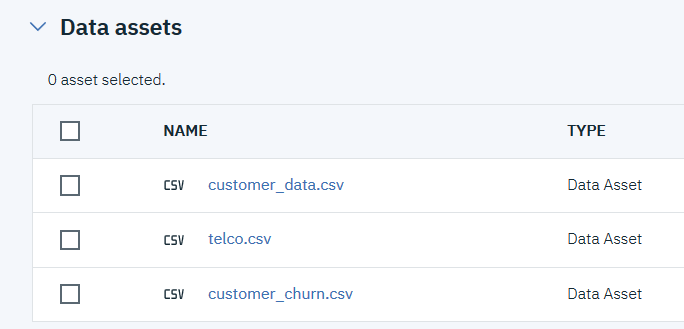




* 1. Haga clic en **Examinar** y cargue todos los archivos .csv desde el directorio **/data** .



* 1. Haga clic en la pestaña **Activos.** Todos los archivos cargados se mostrarán en **Assets** de datos.

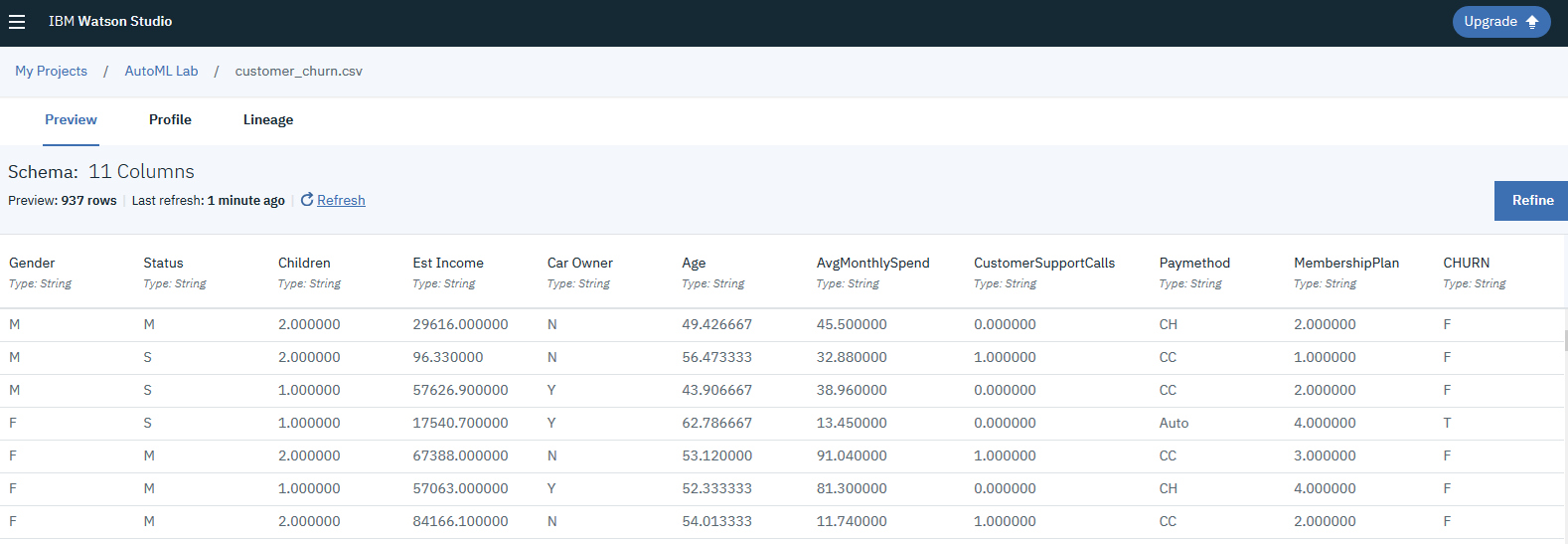


1. Para nuestro primer modelo *de IA automática* usaremos el archivo *customer\_churn.csv* . Haga clic en el archivo para obtener una vista previa de los datos.

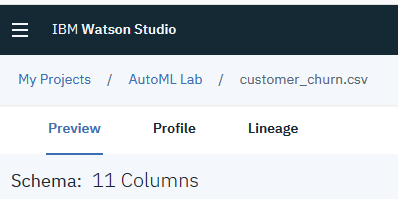
El archivo contiene información de transacciones demográficas e históricas para cada cliente. También contiene la bandera de abandono. Si el valor de marca se establece en *T (True),* significa que la empresa ha sido abandonada.

Con AutoAI, crearemos un modelo que predecirá la renovación para nuevos clientes. Un departamento de marketing puede utilizar esta información para prevenir de forma proactiva el abandono.

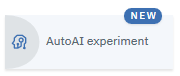
En ciencia de datos, las columnas que usamos para predecir la renovación se denominan *características* y la columna que estamos intentando predecir se denomina *destino*.



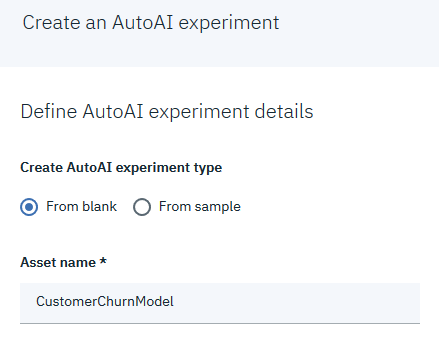
1. Haga clic en **Auto ML Lab** para volver a la vista **Proyecto** .



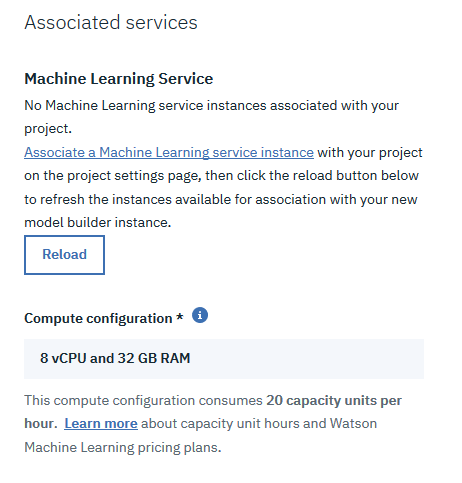
1. En la vista Proyecto**,** seleccione Agregar al proyecto -> Experimento De AutoAI**.**



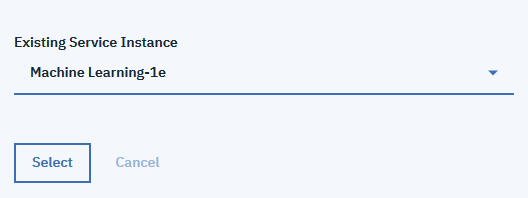
1. Asigne al activo el nombre *CustomerChurnModel.*



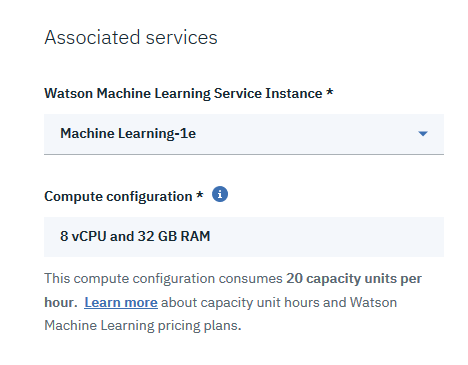
1. A continuación, tenemos que agregar el **archivo WML** a nuestro proyecto.
   1. Haga clic en **Asociar...** enlace.



* 1. Seleccione el **servicio WML**

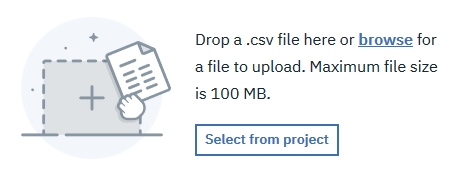


* 1. Haga clic en el botón **Volver a cargar** para asegurarse de que el servicio WML se ha asociado con el proyecto.

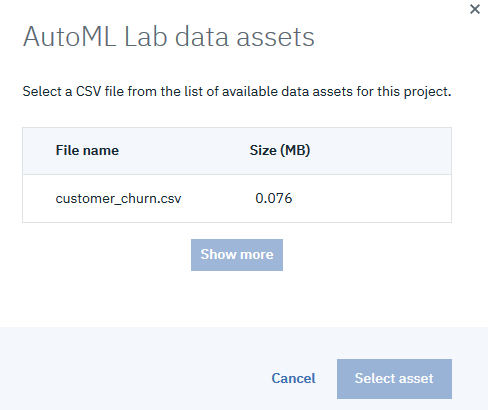


* 1. Haga clic en **Crear** para crear el recurso de AutoAI.

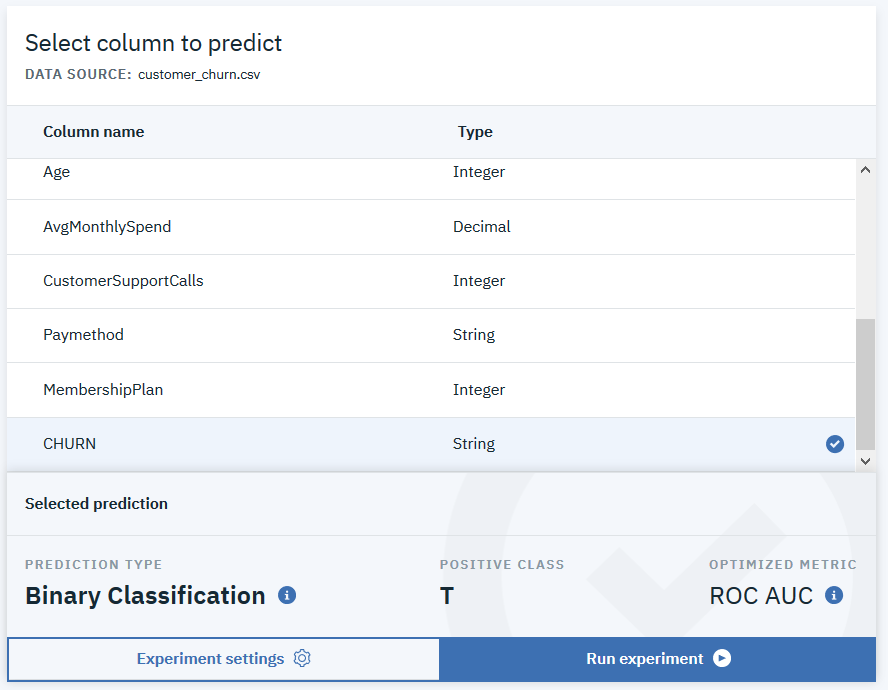
1. En la pantalla Agregar datos de **entrenamiento,** haga clic en Seleccionar **del proyecto**



1. Seleccione *customer\_churn.csv*. A continuación, haga clic en **Seleccionar activo**.



1. En la lista de columnas, desplácese hacia abajo hasta *CHURN* y haga clic en la fila – esto la seleccionará como la variable de destino.



Tenga en cuenta que al hacer clic en la columna para predecir, se seleccionaron varios valores en la parte inferior de la pantalla. Vamos a revisarlos.

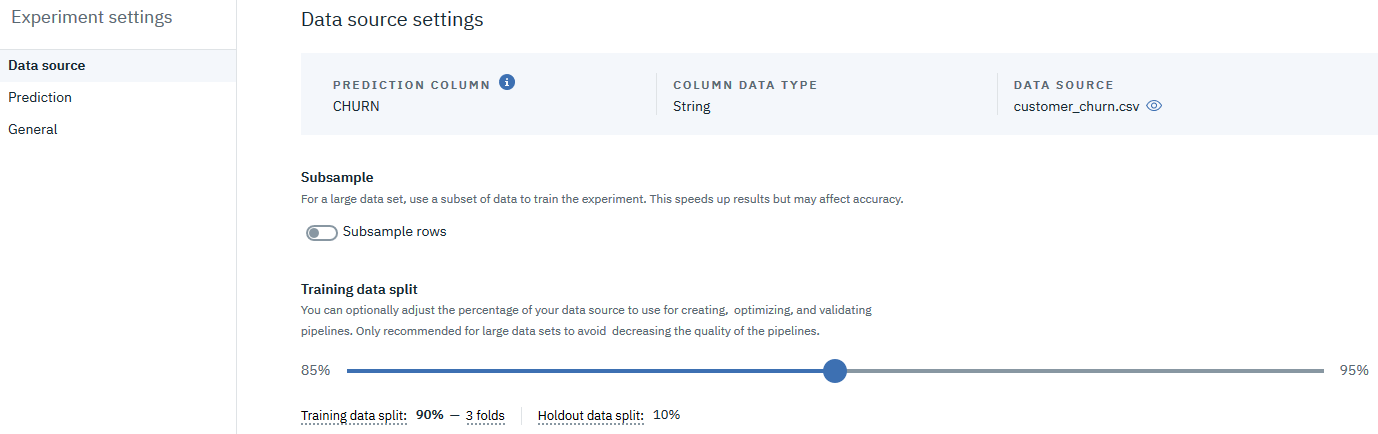
* *Clasificación binaria* es el tipo de clasificación que mejor se adapta a nuestra variable de destino, que tiene 2 *valores,*T (Verdadero) y F (False) y el valor "T" se ha establecido como LA CLASE POSITIVA, que es el grupo de interés.
* *ROC AUC* es una de las métricas de evaluación del modelo. Dado que hay varias métricas de evaluación, necesitamos configurar AutoAI para optimizar el modelo para la métrica específica. Para problemas de clasificación, ROC AUC se utiliza a menudo como la métrica de evaluación primaria.

*Nota: ROC AUC es a menudo preferido sobre la precisión (número total de predicciones correctas) porque mide la capacidad del modelo para predecir "T" como "T" y "F" y "F".*  *En otras palabras, busca precisión para cada clase de predicción.*

Seleccione un nombre de columna diferente, por ejemplo, *AvgMonthlySpend*. Observe que el *tipo de predicción* ha cambiado a *Regresión* y la métrica *Optimizado* ha cambiado a *RMSE*. AutoML cambió estos valores porque *AvgMonthlySpend* es una variable continua.



1. Seleccione CHURN de nuevo y haga clic en **Configuración del experimento**.



Al crear un modelo, los datos se dividen en datos de entrenamiento y pruebas. Los datos de prueba también se denominan "datos de retención". De forma predeterminada, el 10% de los datos se utiliza para las pruebas. Dejemos el valor predeterminado.

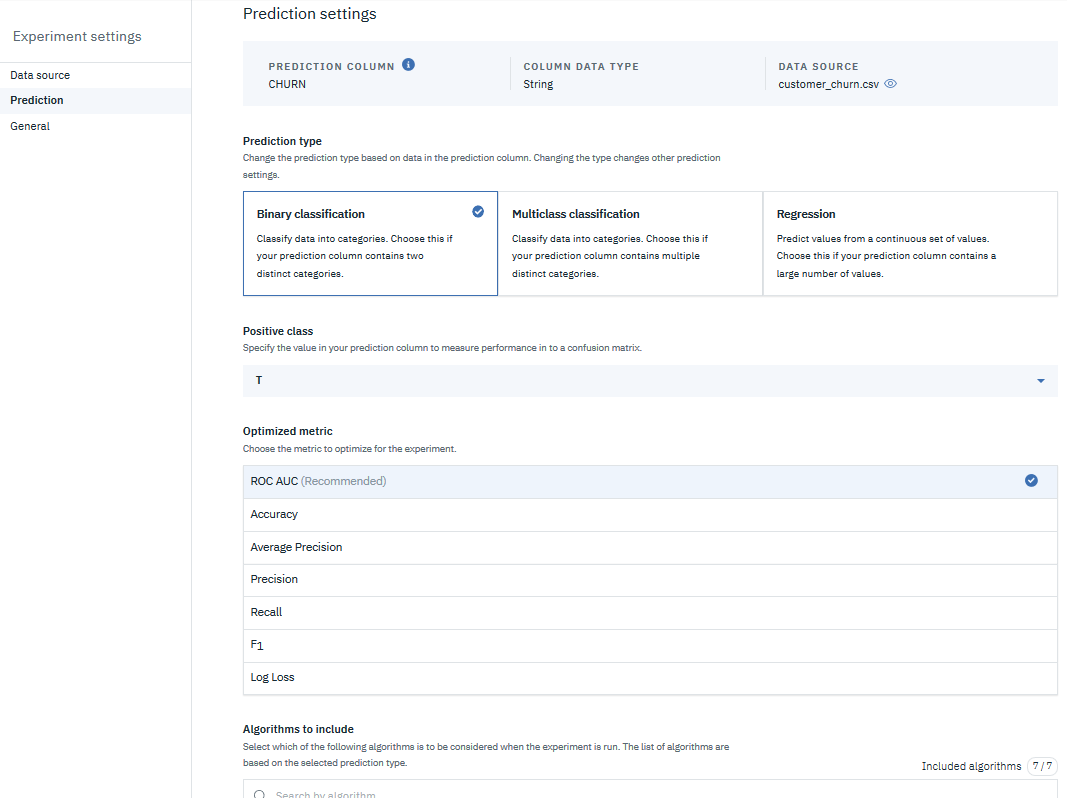
*Los splits* son partes del conjunto de datos de entrada que alternativamente se usarán para el entrenamiento y las pruebas. Al especificar 3 splits, estamos creando 3 partes en el dataset. Usaremos el valor predeterminado de los splits en este laboratorio.

1. Haga clic en la pestaña **Predicción.**

En esta pestaña podemos realizar cambios en el tipo de algoritmo, la clase positiva, la métrica de evaluación del modelo para la que se puede optimizar el modelo y la lista de algoritmos experimentados con.

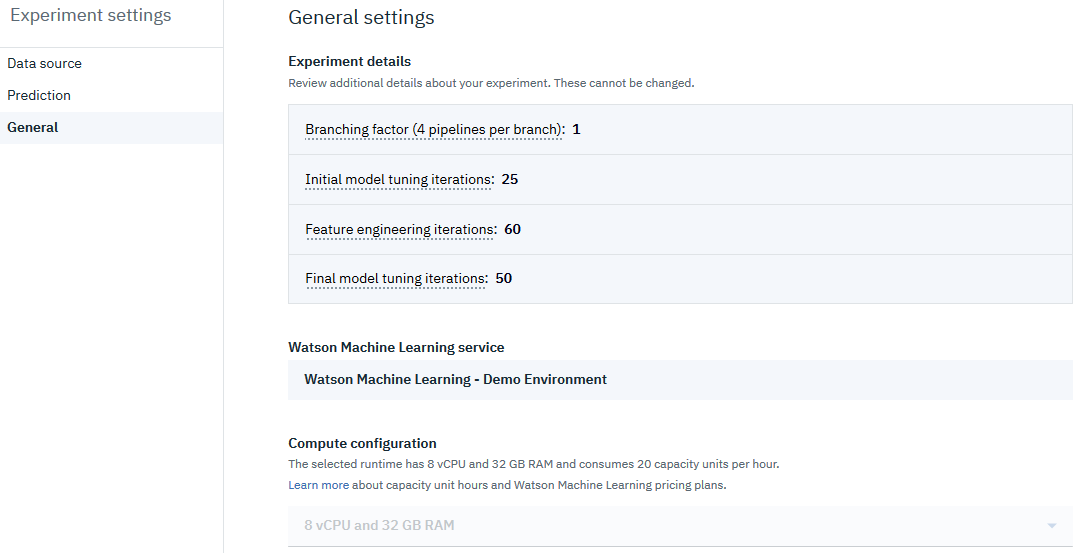
De forma predeterminada, AutoAI ya ha seleccionado los valores que mejor se adaptan al conjunto de datos de entrada y a la variable de destino.

Puesto que no necesitamos realizar ningún cambio, haga clic en **Cancelar**.

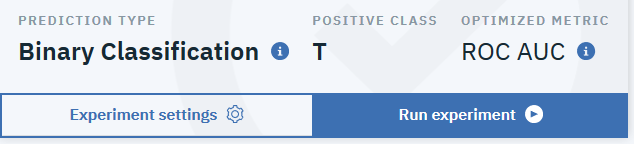


1. Haga clic en la pestaña **General.**

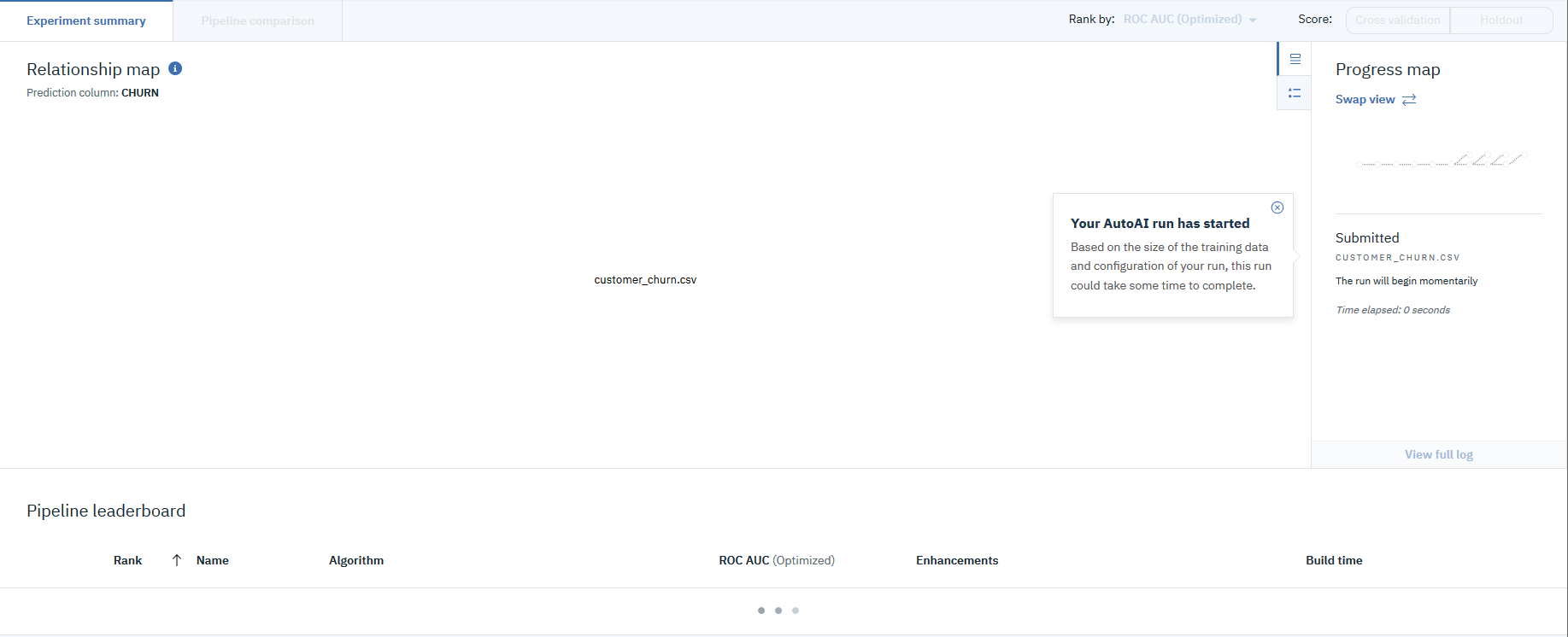
En esta pestaña vemos una lista de la configuración general del modelo que se está aplicando al experimento. No es posible cambiar estos ajustes, pero se muestran para que el usuario pueda ver las condiciones en las que se han creado los modelos.



1. Ahora estamos listos para construir el modelo así que **haga click en cancelar**, ya que no se han realizado cambios
2. Haga clic en **Ejecutar experimento**.

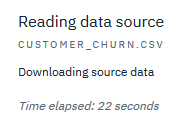


1. *AutoAI* muestra el estado de compilación.



El cuadro de estado en el lado derecho muestra que se envía el trabajo y, a continuación, cambia para indicar las distintas etapas de procesamiento.

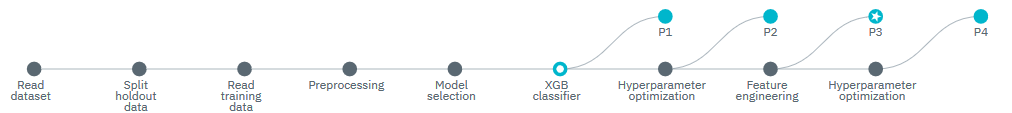
El siguiente estado que se muestra, una vez que el trabajo se ha configurado es leer los datos



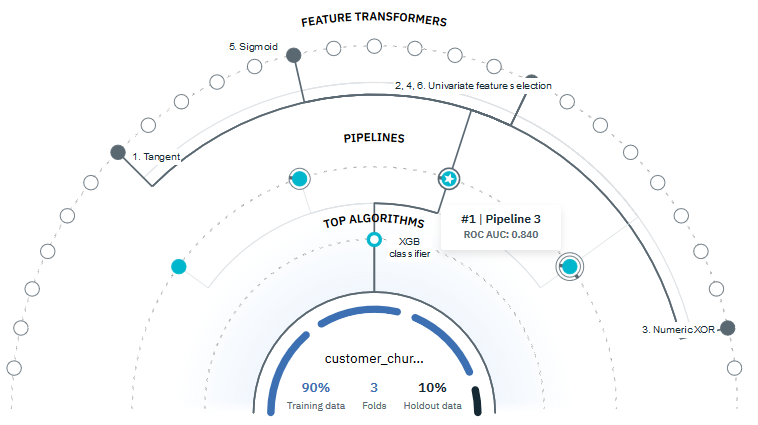
Durante el procesamiento, es posible cambiar entre dos tipos de salida mediante el botón Intercambiar **vista:**

**Mapa de progreso**: muestra la fase actual del procesamiento donde **P1, P2, P3**y **P4** hacen referencia a las 4 pipelines que se construirán. Cada pipeline crea un modelo con diferentes configuraciones.

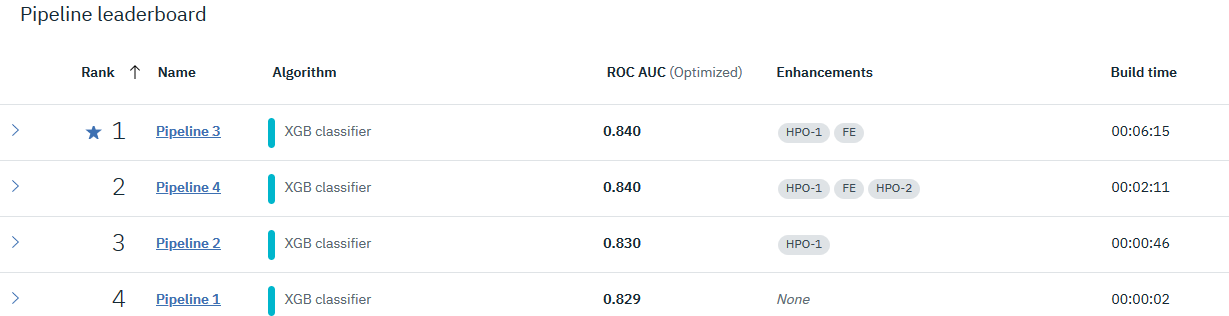
* Como podemos ver en el gráfico, **P1** crea una pipeline utilizando el estimador (algoritmo) que AutoAI ha determinado que es el mejor ajuste para los datos, en este caso **clasificador XGB**.
* **P2 realiza la optimización de** hiperparámetros (HPO). Los hiperparámetros son "configuración" (parámetros) que son específicos de cada algoritmo. Optimización de hiperparámetros significa que estamos construyendo el modelo usando diferentes configuraciones en el algoritmo. *AutoAI* intenta varias combinaciones y determina la combinación que producirá el mejor resultado.
* **P3 realiza ingeniería de** características (deriva nuevas características) y crea un modelo con estas características.
* **P4 realiza la optimización de** parámetros para el modelo que utiliza las entidades derivadas.



**Mapa de relaciones**: muestra qué transformaciones se han aplicado a las distintas pipelines. El gráfico es interactivo, así que coloque el cursor sobre una de las canalizaciones para ver el mapa que se asocia con él



1. Cuando las pipelines se hayan terminado de ejecutar, *AutoAI* mostrará los resultados similares a esta captura de pantalla:



1. Revisemos cada pipeline. Comenzaremos con *Pipeline 1*, que utiliza las características predeterminadas.

Haga clic en *Pipeline 1*.

En la vista **Pipeline** podemos encontrar información diversa sobre el modelo. El lado derecho de la pantalla describe las métricas de evaluaciones que utilizan los científicos de datos.

La **puntuación de validación cruzada** es la puntuación cuando se utilizó la técnica de validación cruzada. La validación cruzada utiliza diferentes partes del mismo conjunto de datos primero para el entrenamiento y, a continuación, para las pruebas.

Un científico de datos experimentado puede examinar una combinación de métricas de evaluación antes de determinar si el modelo está listo para la producción. En un nivel alto, puede utilizar las siguientes pautas para *AUC ROC:*

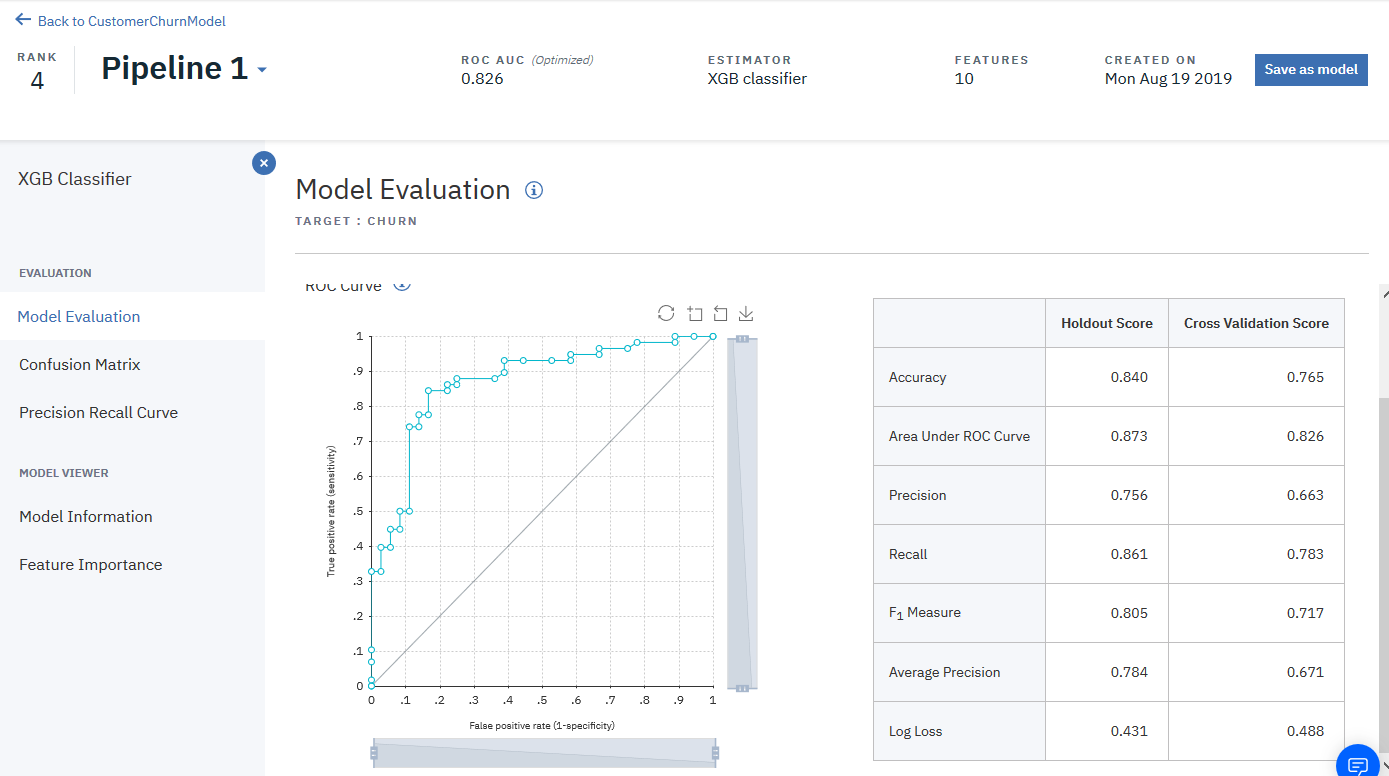
*.90-1 á excelente (A)*

*.80-.90 á bueno (B)*

*.70-.80 á feria (C)*

*.60-.70 á pobres (D)*

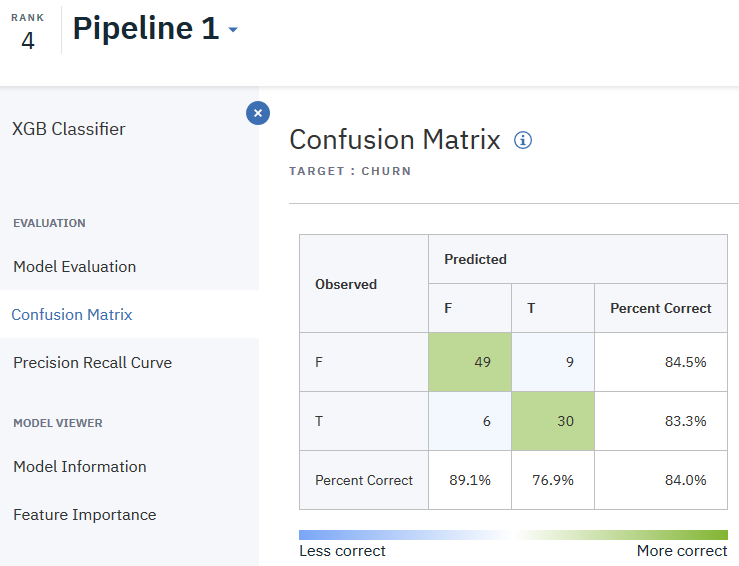
*.50-.60 - fallo (F)*



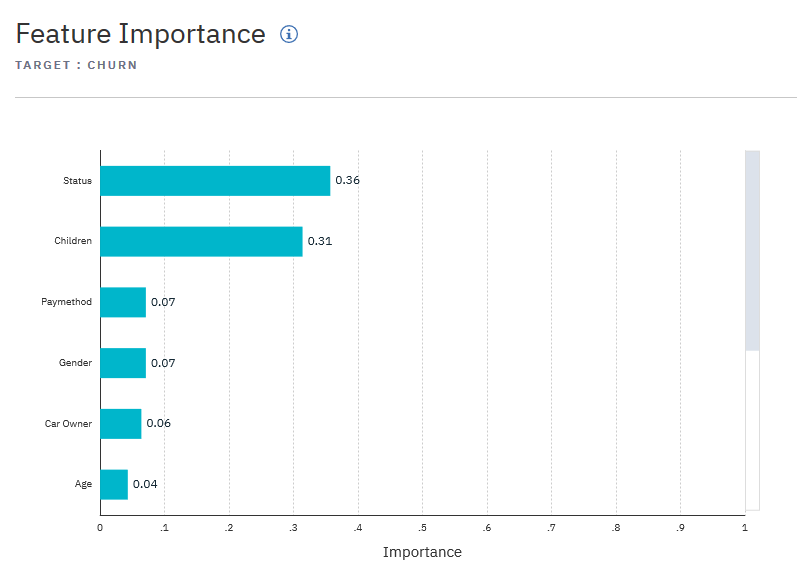
1. Haga clic en **Matriz de confusión** a la izquierda.

La matriz de confusión es otra métrica típica para la evaluación del modelo. Muestra *verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos.*

Por ejemplo, el modelo predijo 49 de F (False – cliente no churn) registros como F (predicción correcta). Al mismo tiempo, predijo 9 registros as T (el cliente va a abandonar), pero el valor real era F. En general, para la predicción False la precisión es del 84%.



1. Click en **Importancia de** la característica. Aquí podemos ver qué características tienen el mayor impacto en la predicción.

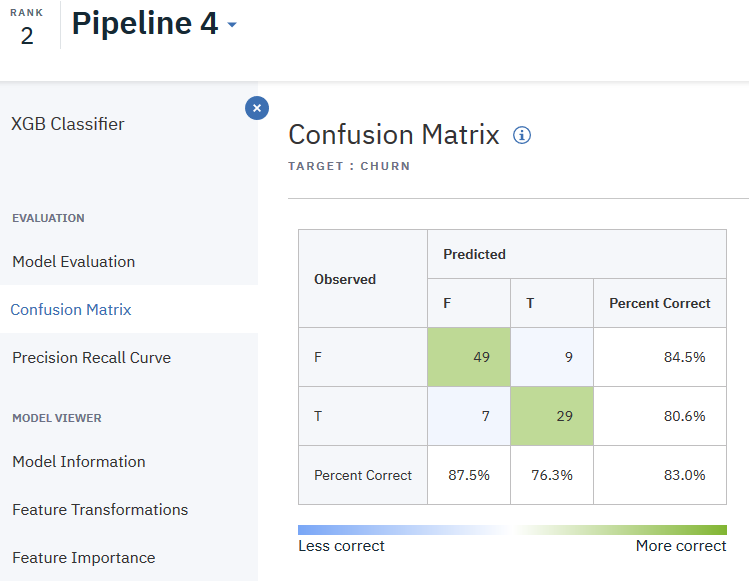


1. A continuación, vuelva a la vista **pipelines** y haga clic en pipeline **4**.

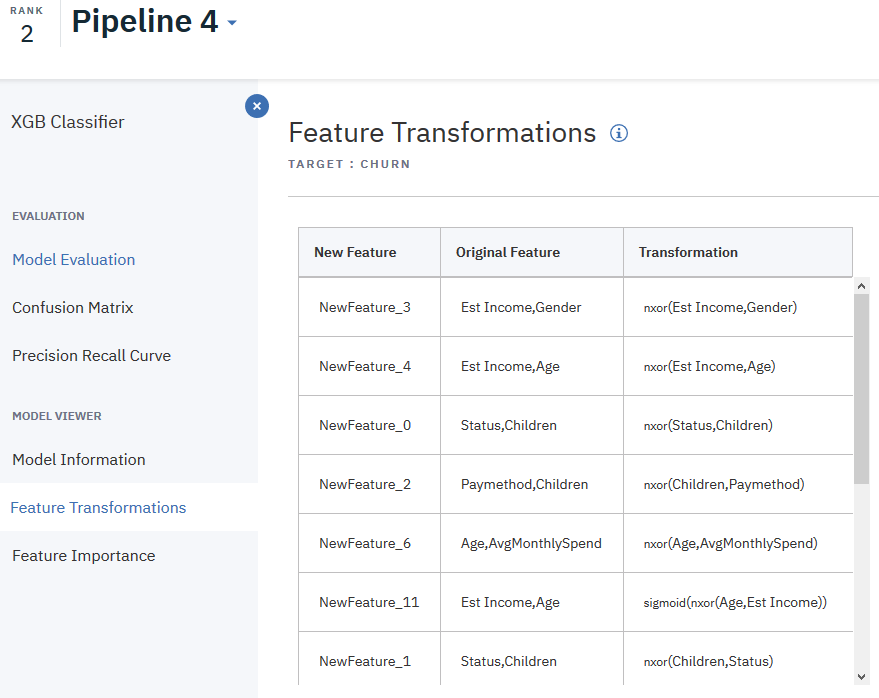
*La pipeline 4* es la tubería para la que se generaron características adicionales y se realizó HPO.

Haga clic en **Matriz de confusión**. Observe que en comparación con *la pipeline* 1, *la canalización 4* tiene diferentes porcentajes de predicciones correctas para T y F.

Un científico de datos puede utilizar esta información para seleccionar el mejor modelo. Por ejemplo, en un caso de uso de fraude, una tasa de falso sin negativo (el fraude no fue predicho por el modelo) puede ser tan importante como la tasa positiva real.

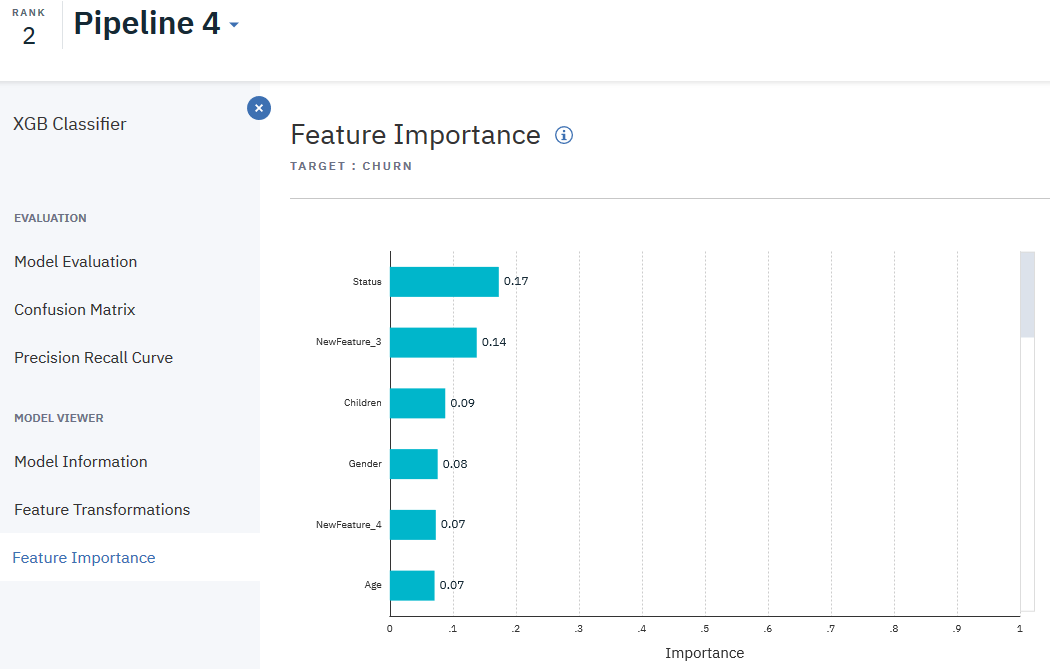


En *la canalización 4* tenemos características adicionales generadas por el modelo. Puede revisarlos seleccionando **Transformación de características**.



Si desea comprender por qué AutoAI realizó transformaciones específicas, puede revisar la documentación: <https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-details.html?audience=wdp#data-transformations>

Si hace clic en **Importancia** de entidad, observará que una de las características generadas se convirtió en una de las características clave para predecir la renovación.

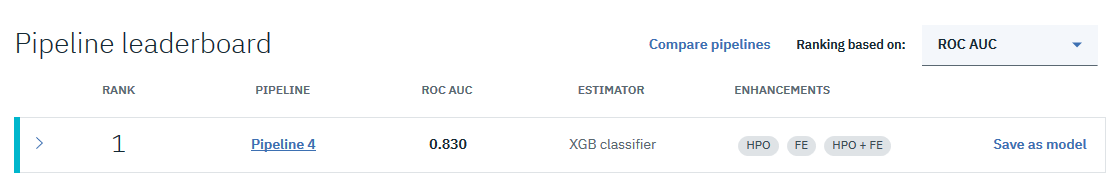


1. Si lo desea, revise la salida para **pipelines 2** y **3**.

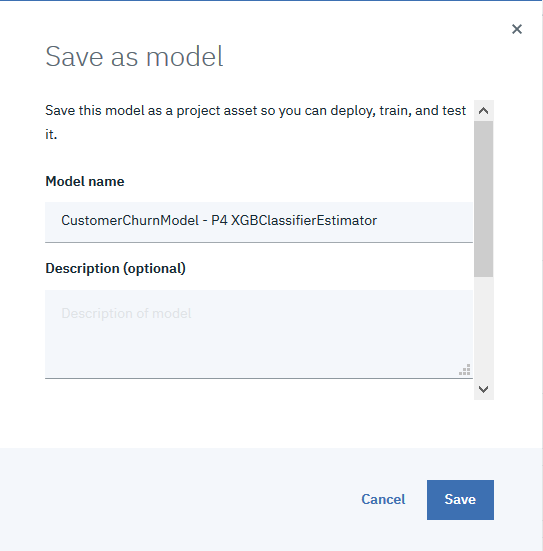
En nuestro ejemplo no hay una gran diferencia en el rendimiento de cada canalización, pero dado que Pipeline 4 tiene el porcentaje más alto de positivos verdaderos (predicción de abandono correcta), lo seleccionaremos para la implementación.

*Nota: los resultados pueden ser ligeramente diferentes porque la selección de pliegues es diferente para cada proceso de construcción del modelo. Puede seleccionar* *la canalización con el mejor rendimiento para la implementación.*

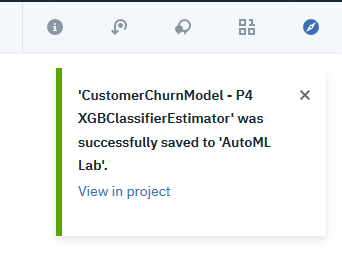
Haga clic en **Guardar como modelo** junto a **pipeline 4**.



Utilice el nombre del modelo predeterminado y haga clic en **Guardar**.

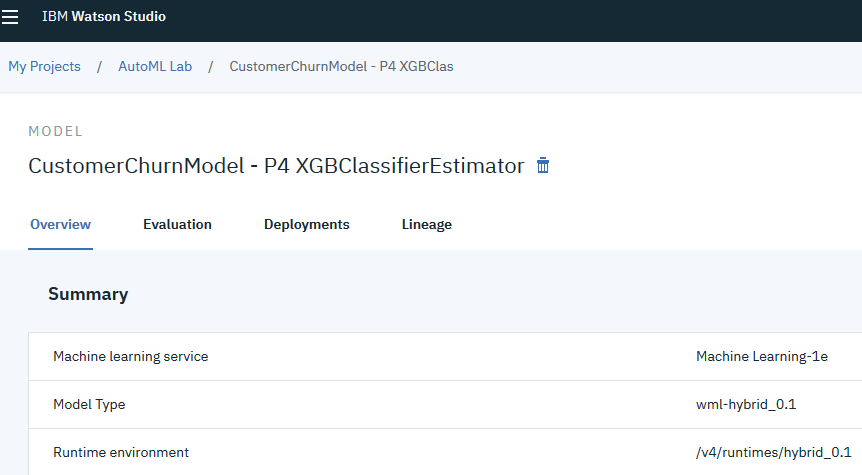


El estado de guardado se muestra en la esquina superior derecha de la pantalla. Haga clic en el vínculo **Ver en el proyecto.**



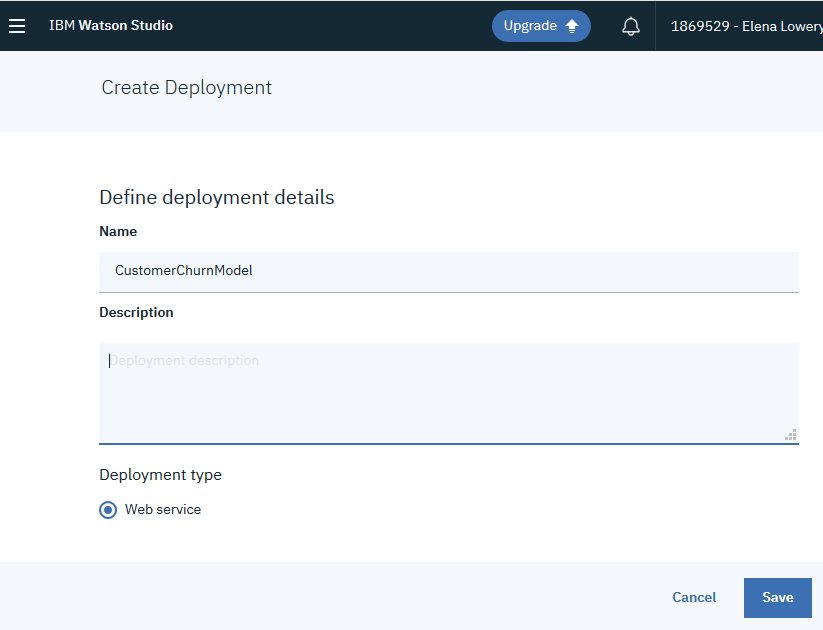
1. A continuación, configuraremos la implementación.

Al hacer clic en el vínculo del paso anterior, se obtiene la vista de modelo.

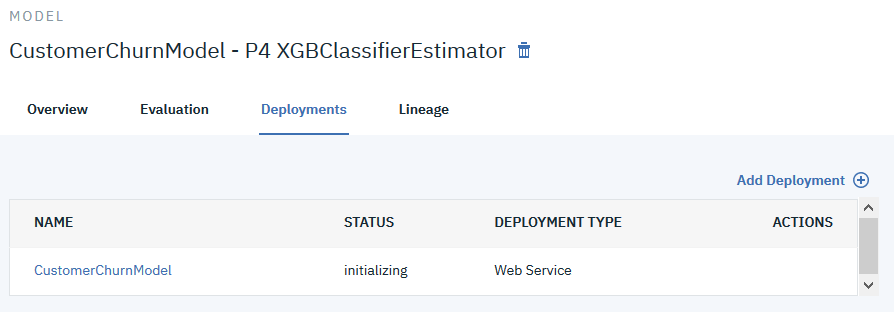


Haga clic en la pestaña **Implementaciones** y, a continuación, en Agregar **implementación**.

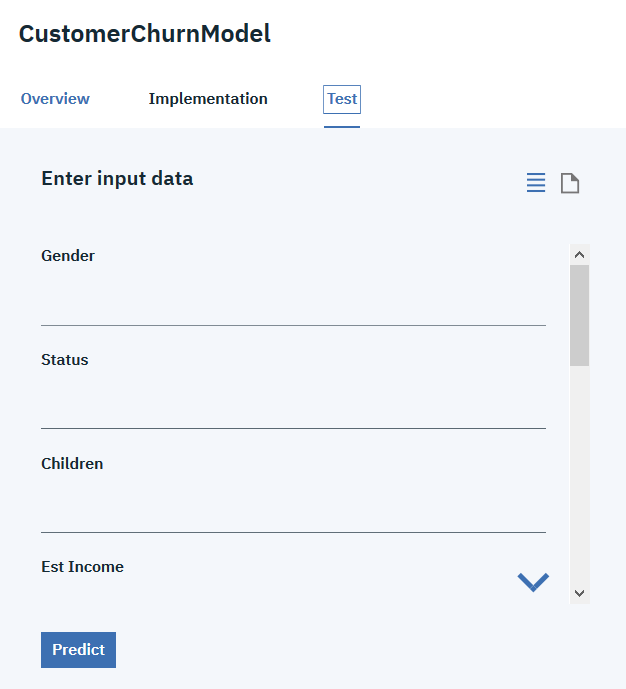
Proporcione el nombre del deployment *CustomerChurnModel*y haga clic en **Guardar**.



Espere hasta que el estado cambie de *inicializando* a *listo*.



1. Haga clic en la implementación (*CustomerChurnModel*) y, a continuación, haga clic en la ficha **Probar** en la vista de detalles del modelo.



Introduzca los siguientes valores para las pruebas.

**Género**: F

**Estado**:M

**Niños**: 2

**Ingresos**Est : 45430

**Propietario del coche:** Y

**Edad**: 61 años

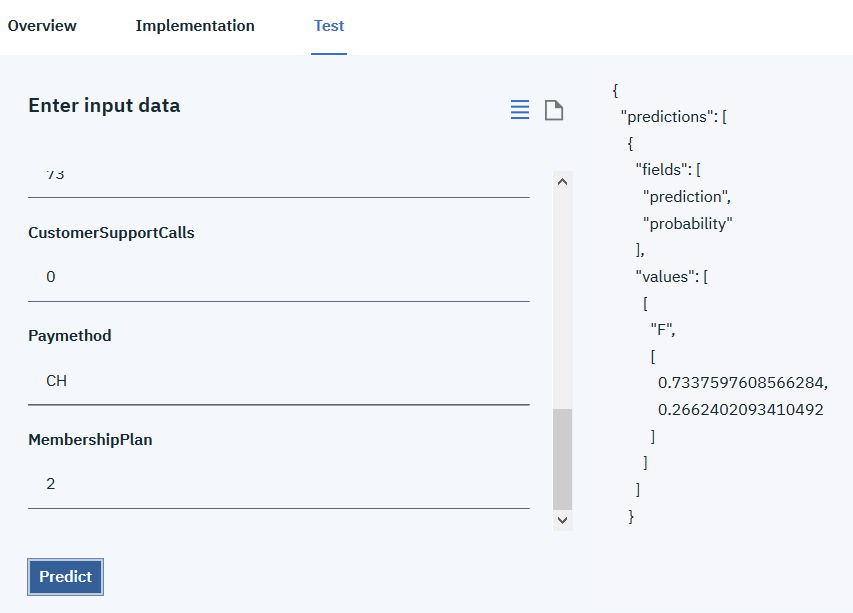
**Gasto Smonthly**: 73

**CustomerSupportCalls**: 0

**Método de pago**: CH

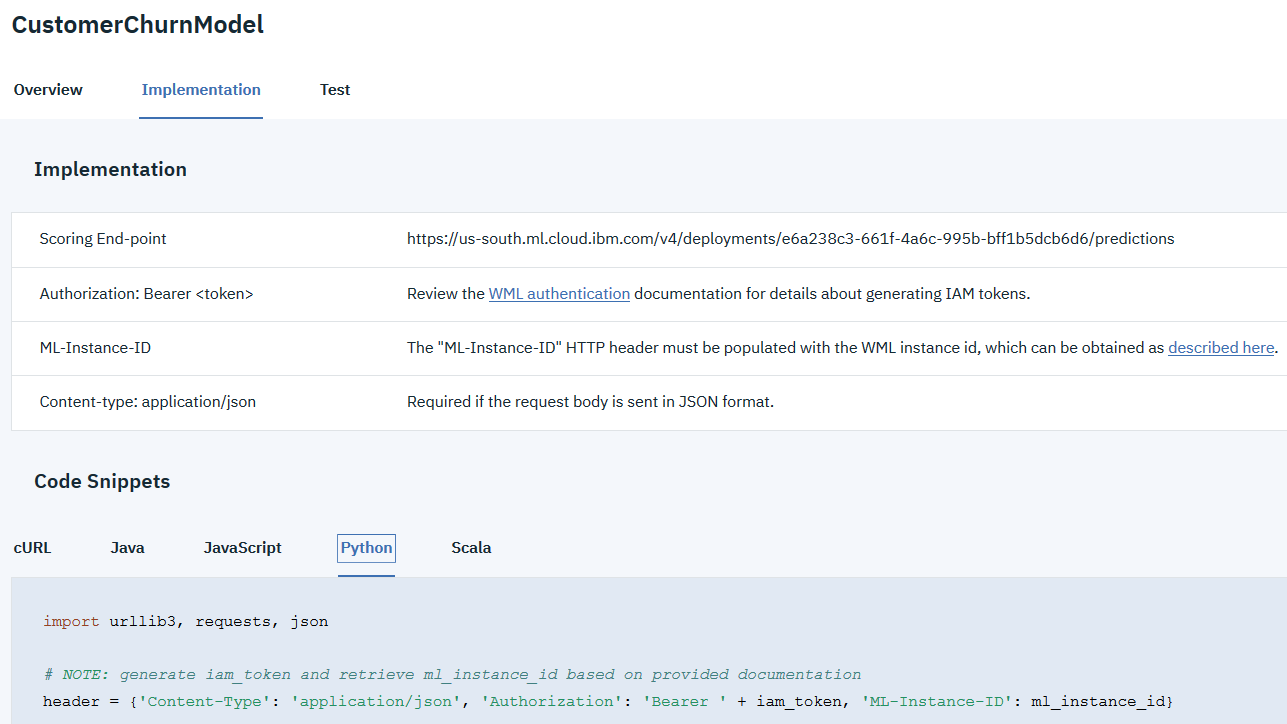
**Plan de Membresía**: 2

El modelo devuelve F (false): no es probable que el cliente se desprenda de nuestros serivicios con una confianza del 73 por ciento.



Ha terminado de desarrollar, implementar y probar un modelo de AutoAI. El modelo se puede integrar con otras aplicaciones mediante la API de REST.

Puede encontrar ejemplos de código para invocar el modelo en la ficha **Implementación.**



# AutoAI en SPSS Modeler

Al igual que las capacidades de AutoAI que se tratan en la última sección, **Modeler** automatiza el proceso de extremo a extremo de desarrollo e implementación de modelos. Algunas diferencias son:

* *Un enfoque más granular (paso a*p aso): un científico de datos puede tomar decisiones/modificaciones.
* *Implementación diferente*: Modeler utiliza la implementación de IBM, que se documenta en la Guía de algoritmos de SPSS Modeler: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.2/en/AlgorithmsGuide.pdf>

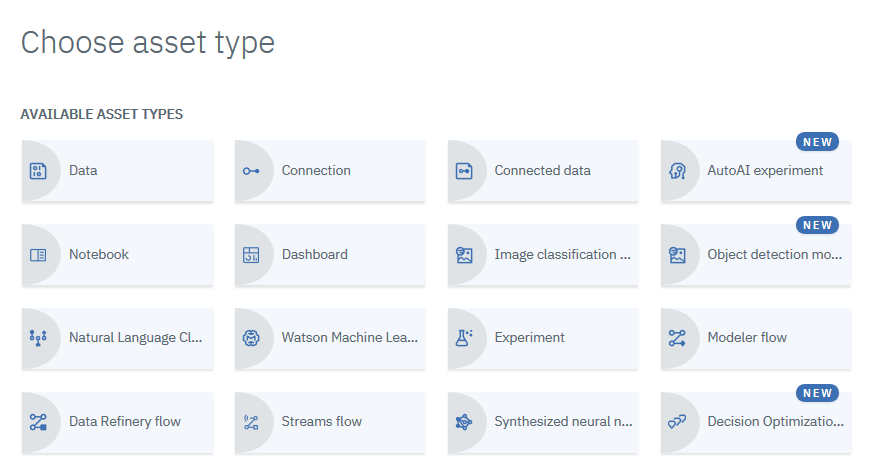
SPSS Modeler contiene varias características de AutoAI:

* Modelado automático para modelos de clasificación, regresión y agrupación en clústeres
* Preparación automática de datos
* Auditoría de datos
* Selección automática de funciones
* Características automáticas en algoritmos individuales (por ejemplo, regresión).

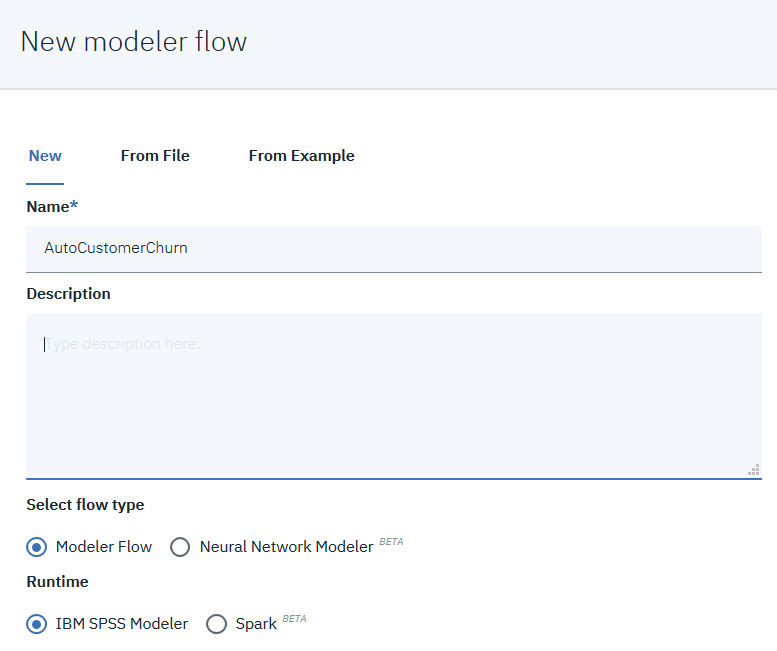
En este laboratorio nos centraremos en 3 características: modelado automático (caso de uso de clasificación), selección de características y preparación automática ded atos.

## Modelado automático

1. En **Watson Studio Cloud,** vaya a la vista **Proyecto.** Seleccione **Agregar al proyecto -> Modeler Flow**

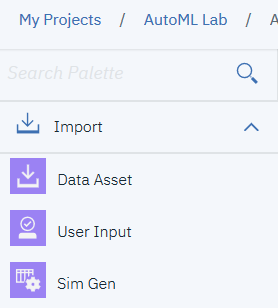


1. Introduzca el nombre de flujo *AutoCustomerChurn*. Mantenga los valores predeterminados para el tipo de **flujo** y el tiempo de **ejecución.**

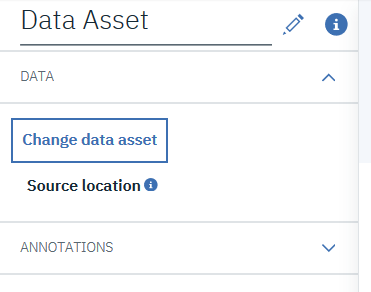


Para el caso de uso de clasificación, se utilizará el mismo conjunto de datos que hicimos en la sección anterior.

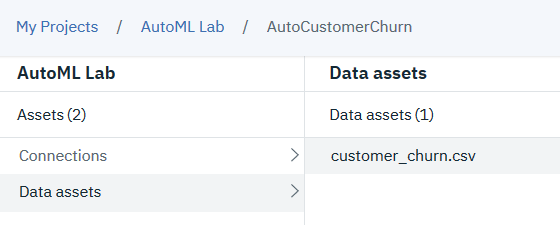
1. Expanda la pestaña **Importar.** Seleccione **Data Asset** y arrástrelo al lienzo.



1. Haga doble clic en el icono **Data Asset**. En la vista **Propiedades**, haga clic en cambiar **Data Asset.**

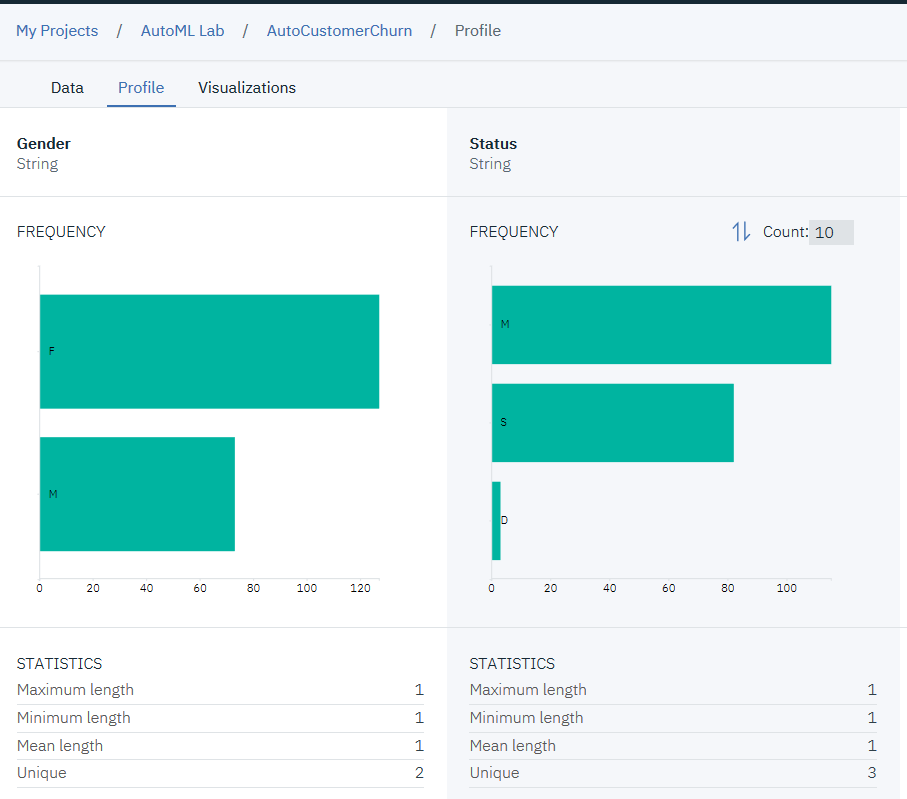


1. Expanda **Data Assets** y seleccione *customer\_churn.csv*. Haga clic en **Aceptar**.



1. Haga clic con el botón derecho en el nodo **Data Asset** y seleccione **Perfil**. La pantalla **Perfil** proporciona la vista previa del conjunto de datos, las estadísticas de resumen (en la página **Perfil)** y la interfaz para crear visualizaciones.

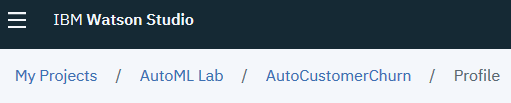
*Nota: si recibe el error ElementLocked, haga clic en Aceptar.*



A continuación, agregaremos los nodos **Type**, **Partition** y **AutoClassifier**

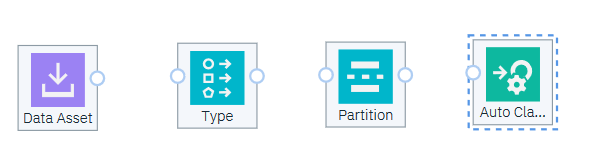
* En el nodo **Tipo** especificaremos las entidades y las variables de destino. También nos aseguraremos de que Modeler determinara correctamente los tipos de datos.
* En el nodo **Partición** especificaremos la división entre los datos de entrenamiento y de prueba.
* El nodo **AutoClassifier** contiene un conjunto de algoritmos que se usarán para entrenar el modelo.

1. Vuelva al lienzo seleccionando el flujo en el menú de navegación.



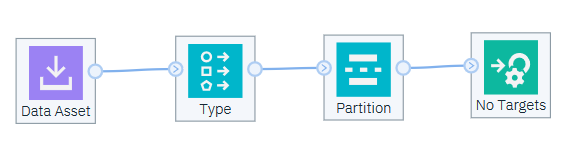
1. En la pestaña **Operaciones** de campo, arrastre los nodos **Tipo** y **Partición** al lienzo. A continuación, arrastre el nodo **AutoClassifier** (desde la pestaña **Modelado)** al lienzo.

Cuando haya terminado, la transmisión debería ser similar a esta captura de pantalla:



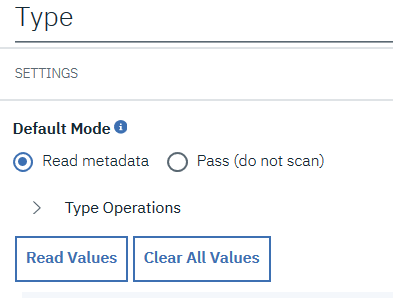
Conectar todos los nodos de izquierda a derecha: los datos "fluirán" a través de los nodos y las operaciones se aplicarán en la dirección del flujo.

Las conexiones se realizan arrastrando el ratón desde/hacia los puntos de conexión. Cuando haya terminado, el flujo debería ser similar a esta captura de pantalla.

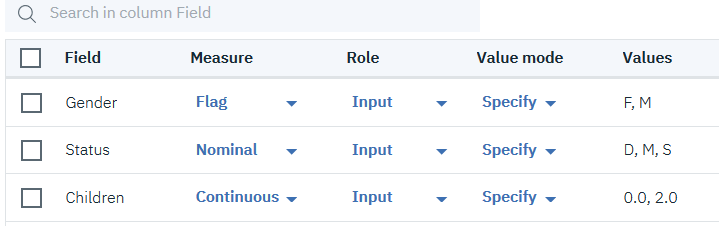


A continuación, necesitamos configurar los nodos.

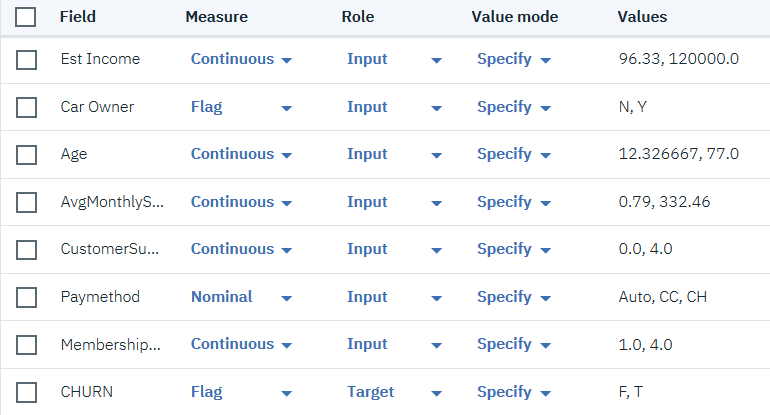
1. Haga doble clic en el nodo **Tipo.** En las propiedades del nodo **Tipo,** haga clic en **Leer valores**.



Leer los valores "inicializa" el **tipo** nodo. Observe que ahora muestra valores para cada campo. El nodo de tipo también ha determinado el tipo de variable *(Flag, Nominal, Continuous,*etc.)

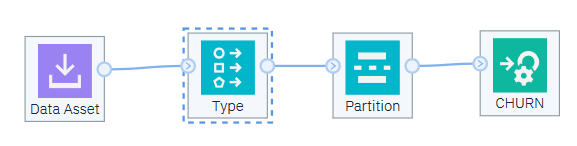


Necesitamos realizar 2 cambios en el nodo **Tipo:** establezca el **rol** de *CHURN* en *Target* ,­cambie **Medir** a *indicador*.

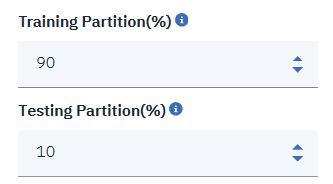


Haga clic en **Guardar**.

Observe que el nodo **AutoClassifier** ahora conoce la variable de destino: el icono ha cambiado para mostrar CHURN.

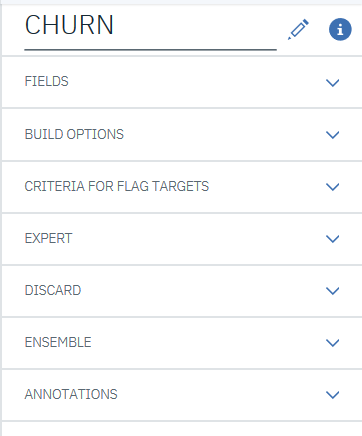


1. Haga doble clic en el nodo **Partición.** Cambie la división entre los datos de entrenamiento y prueba a *90* y *10*. Haga clic en **Guardar**.



1. No es necesario realizar ningún cambio en la configuración del nodo **AutoClassifer,** pero si desea revisarlo, haga doble clic en el nodo.

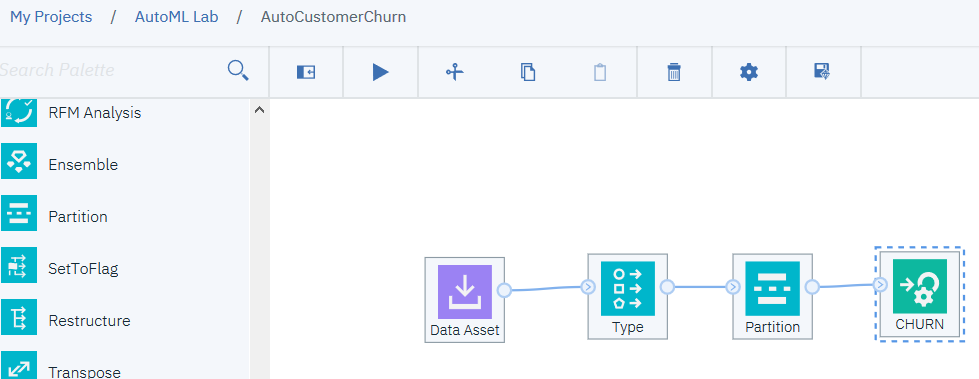
Un científico de datos experimentado puede modificar varias opciones en la configuración del nodo.



Usaremos la configuración predeterminada para el resto de los valores.

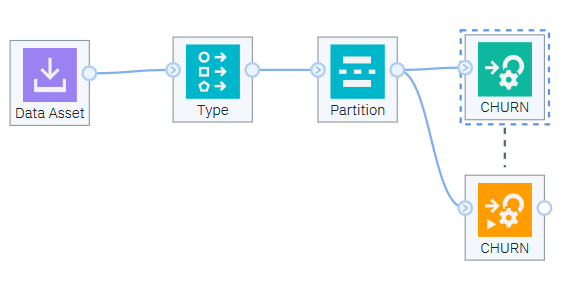
Haga clic en **Cancelar** para volver al lienzo.

1. Ahora estamos listos para construir el modelo. Haga clic en el icono **Ejecutar.**



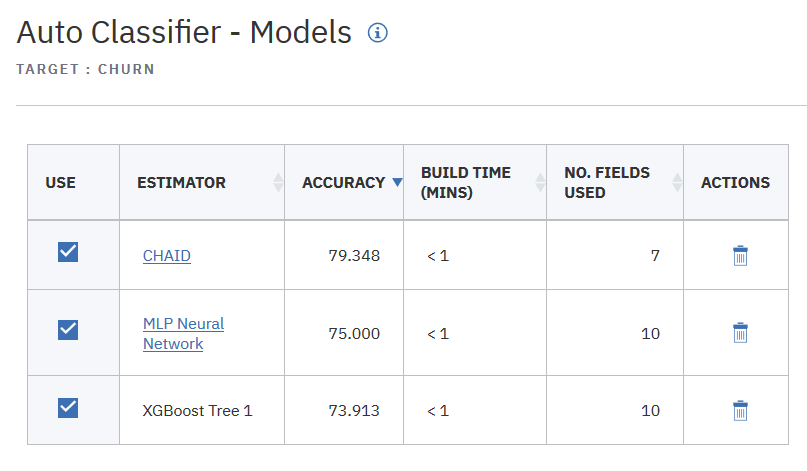
1. Cuando se haya creado el modelo, aparecerá un icono naranja en el lienzo.

*Nota: En Modeler los nodos son color-codificado por función. Como se puede ver en el flujo que hemos construido, los iconos púrpuras son fuentes de datos, azulado – manipulación de datos, verde – algoritmos y naranja – modelos.*



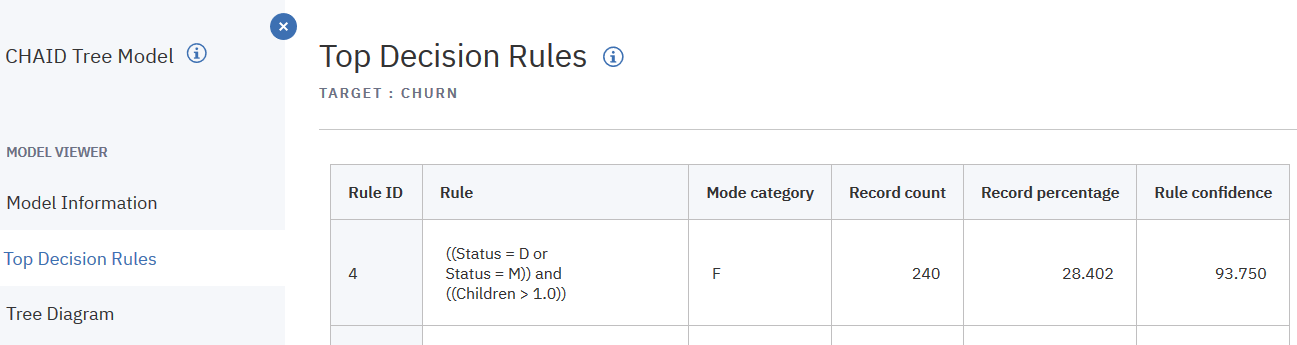
1. Haga clic con el botón derecho en el modelo generado y seleccione **Ver modelo**.

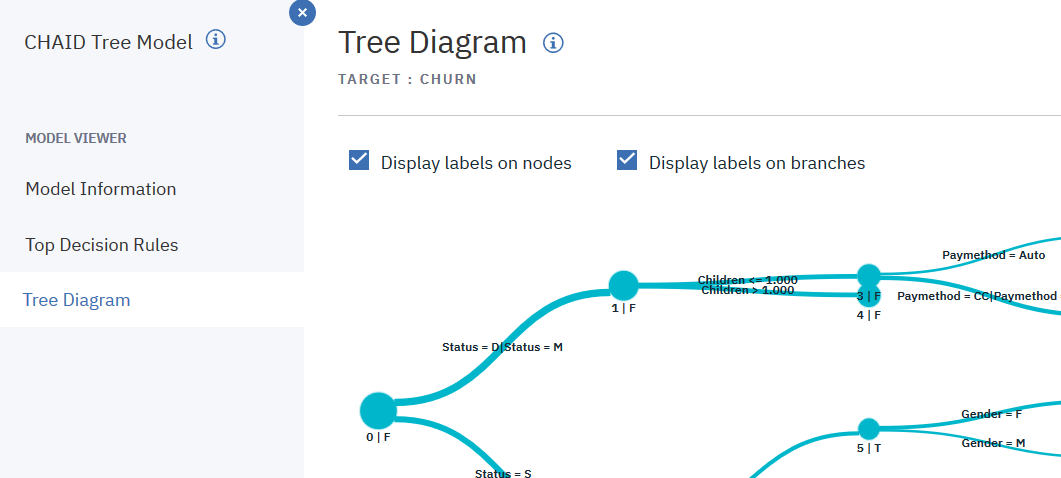
Modeler mostró los 3 modelos principales.



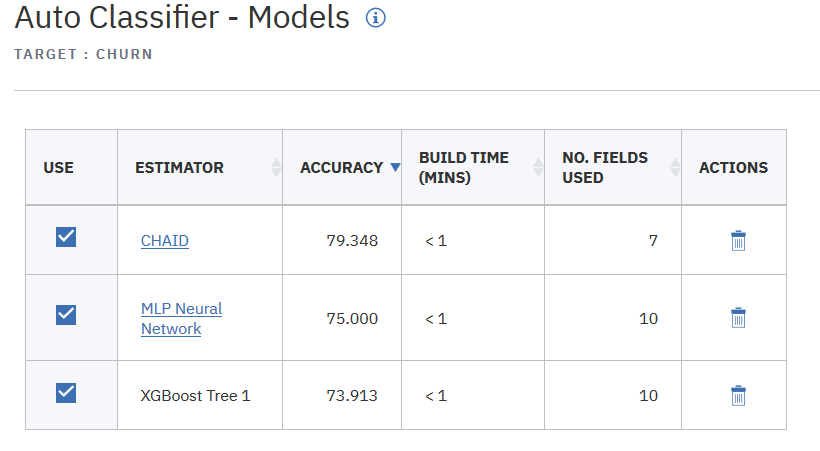
1. Podemos profundizar en los modelos *CHAID* (modelo de alto rendimiento) y *MLP Neural Network.*

Haga clic en **CHAID** y revise las reglas de **decisión superior** y el diagrama de **árbol**.





Vuelva a la vista de **resumen Modelo.**



En este punto un científico de datos puede tomar algunas acciones. Si mantenemos la configuración predeterminada (las casillas de verificación para todos los modelos están marcadas – Modeler realizará la puntuación de conjunto. Los detalles de las ponderaciones se proporcionan en la documentación del modelador.

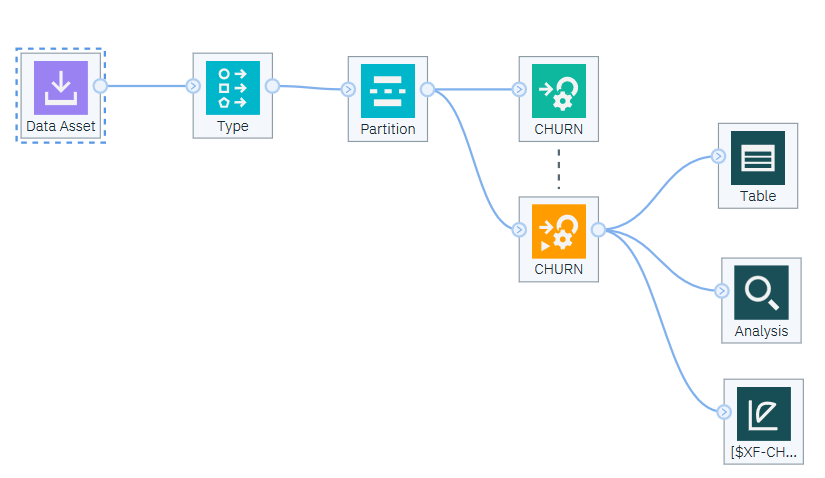
Un científico de datos puede eliminar o desmarcar los modelos del conjunto (no se utilizarán para puntuar).

Por último, un científico de datos puede utilizar el modelado automático como un paso rápido para determinar qué modelo se debe utilizar. Más adelante pueden crear un modelo individual para el rendimiento superior (todos los algoritmos que se usan en nodos automáticos están disponibles como algoritmos individuales en la pestaña Modelado).

En nuestro laboratorio usaremos la configuración predeterminada.

*Nota: En AutoClassifier los modelos se clasifican en función de la precisión, pero Modeler también proporciona opciones más detalladas para la evaluación del modelo (por ejemplo, Matriz de confusión).*

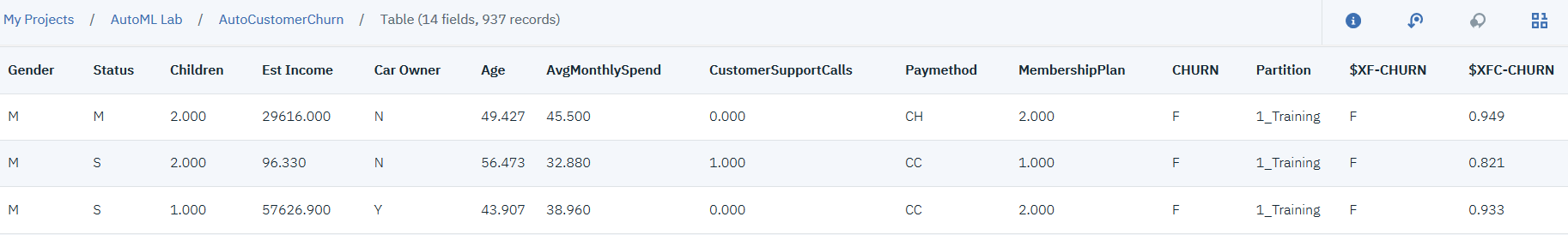
1. Vuelva al lienzo. Agregar **tabla** (ficha **Salida),** **Análisis** (ficha **Salida)** y **Evaluación** (ficha **Gráficos).**



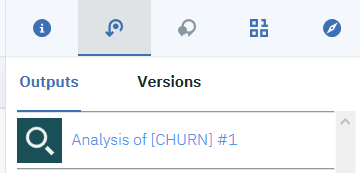
No es necesario configurar ninguno de estos nodos. Ejecutaremos cada nodo individualmente seleccionando **Ejecutar** en el menú contextual del nodo.

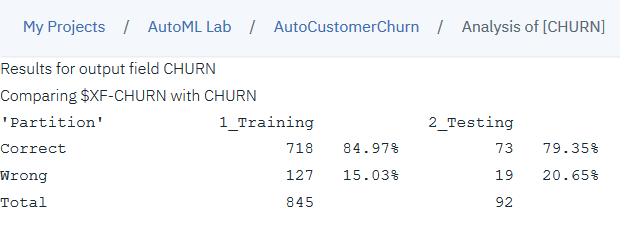
1. Ejecute el nodo **Tabla.**

Cuando ejecutamos el nodo **Tabla,** puntúa los datos y los muestra en un formato de tabla. Las dos últimas columnas, *$XF-CHURN* y *$XFC-CHURN* son los resultados de puntuación: la predicción y la confianza en la predicción.

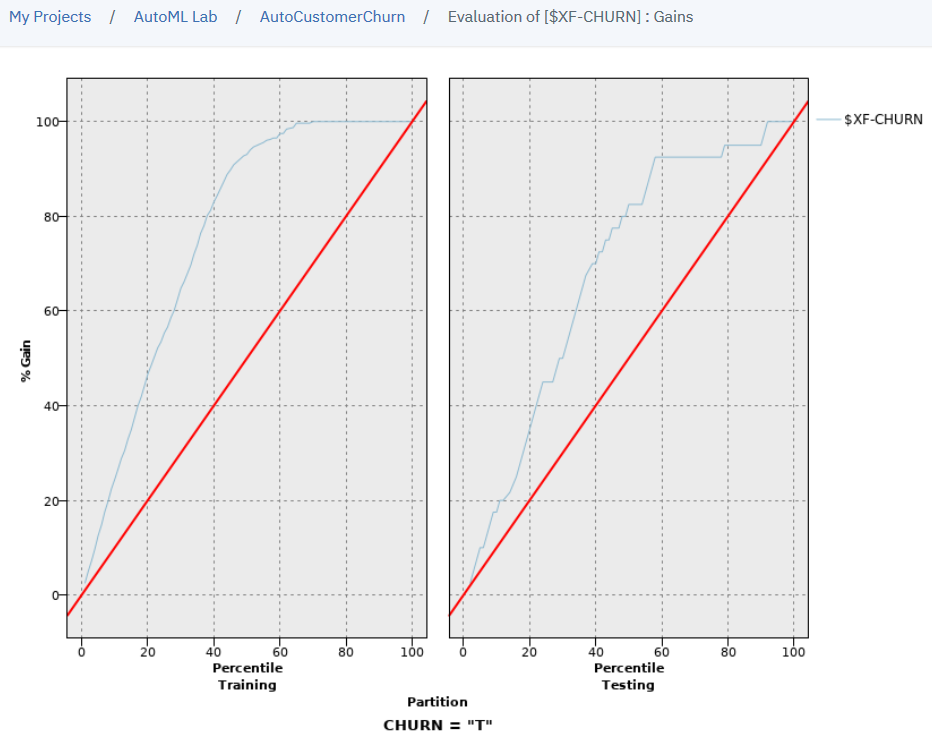


1. Ejecute el nodo **Análisis.** Haga doble clic en él en el panel **Salida** para ver la **Precisión**.





1. Ejecute el nodo **Evaluación.** Haga doble clic en él en el panel **Salida** para ver la **Evaluación**.



Como ha visto, los flujos SPSS son similares a las pipelines en código abierto: contienen varios pasos.

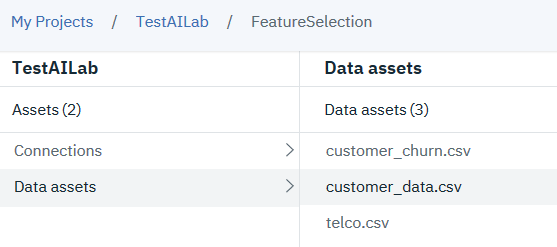
## Selección de características

La selección de características en Modeler se puede utilizar como una tarea independiente o en combinación con la preparación y el modelado de datos.

Hemos creado un "flujo de inicio" para esta sección.

1. Vaya al proyecto y seleccione Agregar flujo **de modelador.**
2. Seleccione **Desde archivo**. Vaya al directorio */flows* de los archivos de laboratorio descomprimidos e importe la secuencia *FeatureSelection.*
3. Si se le pide que cambie el **Data Asset** ,cargue el archivo *customer\_data.csv* desde el directorio */data.*

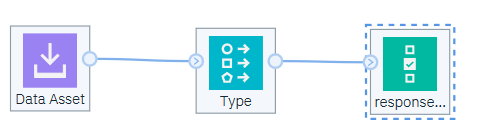
Si no ha visto el mensaje, haga doble clic en **Data Asset** , haga click en **Cambiar** Data Asset seleccione *customer\_data.csv* en el proyecto.



1. La secuencia *FeatureSelection* tiene solo 2 nodos: el **recurso** de datos y que el **tipo** nodo. Hemos preconfigurado el nodo **Tipo** con la medida correcta y la variable de destino.

Si lo desea, obtenga una vista previa de los datos.

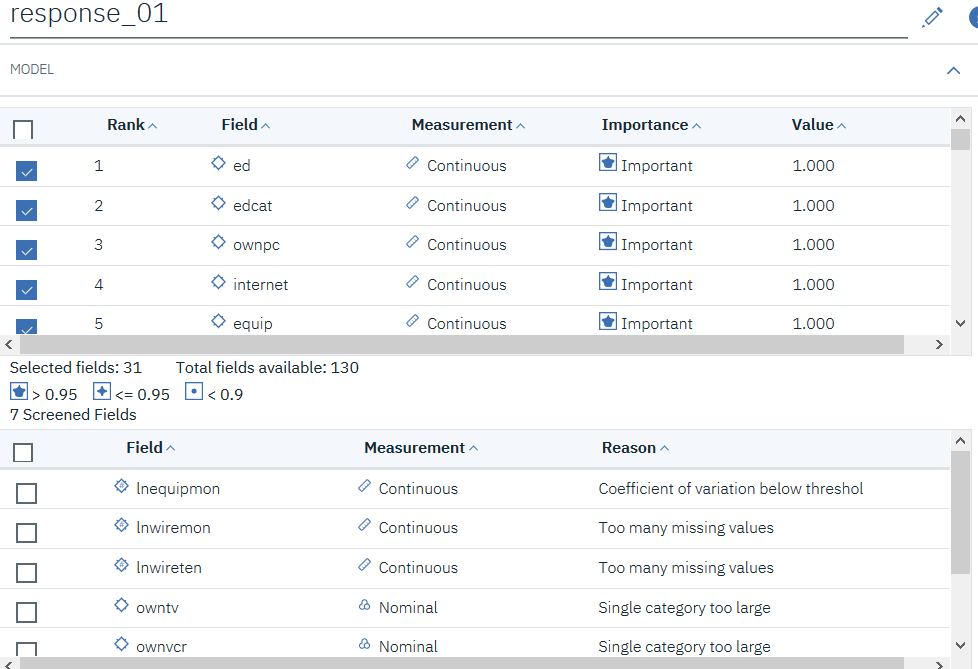
1. Conecte el nodo **Selección** de características (desde la ficha **Modelado)** al nodo **Tipo.**



1. Haga clic con el botón derecho en el nodo **Selección** de características y seleccione **Ejecutar**.
2. Haga doble clic en el modelo generado y revise los resultados.

Revise el valor **importancia** de cada campo. Los campos que son de mercado como *Importante* se seleccionarán automáticamente para el resto del flujo.

La importancia de los campos "pantallados" no se pudo determinar debido a los problemas mostrados en la salida.



1. Agregue un nodo **Tipo** después del nodo de **Selección** de características. Haga doble clic en el nodo **Tipo.**

Observe que se han descartado todos los campos que no son importantes.

Como se mencionó anteriormente, un científico de datos puede seguir creando este flujo de Modeler o simplemente utilizar los resultados del análisis para otros fines (por ejemplo, la implementación del modelo en código abierto).

## Preparación automática de datos

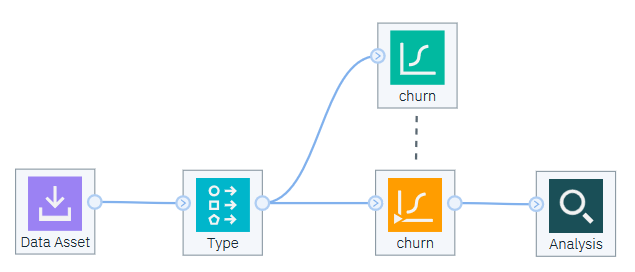
La preparación automática de datos en Modeler se puede utilizar como una tarea independiente o en combinación con la selección y el modelado de características.

Hemos creado un "flujo de inicio" para esta sección.

1. Vaya al proyecto y seleccione Agregar flujo **de modelador.**
2. Seleccione **Desde archivo**. Vaya al directorio */flows* de los archivos de laboratorio descomprimidos e importe la secuencia *AutoDataPrep.*
3. Cuando se le pida que cambie el **Data Assets**, cargue el telco.*csv* de *la compañía telefónica* desde el directorio */data.*

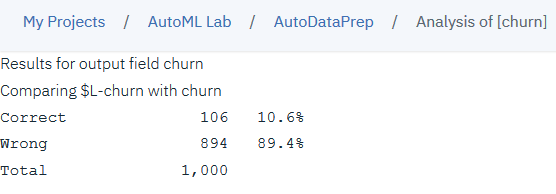
Si no ha visto el mensaje, haga doble clic en **Asset Data Asset**(DataAsset), haga clic en Change Data Asset(Cambiar activo de **datos)**y seleccione telco.csv en el proyecto.

1. La secuencia *AutoDataPrep* incluye un modelo de regresión logística para predecir la rotación del cliente. Vamos a ejecutar el modelo y revisar la precisión. Haga clic con el botón derecho en el nodo **Análisis** y seleccione **Ejecutar**.

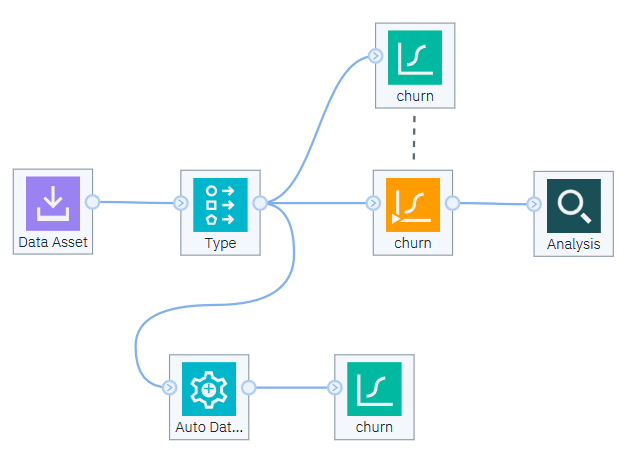


1. Revise la precisión haciendo doble clic en la salida **Análisis** en el panel **Salida.**

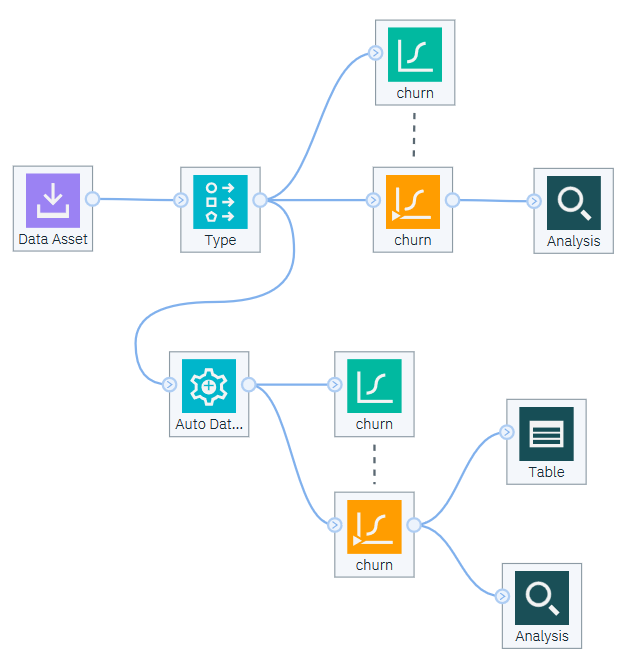
La precisión es pobre.



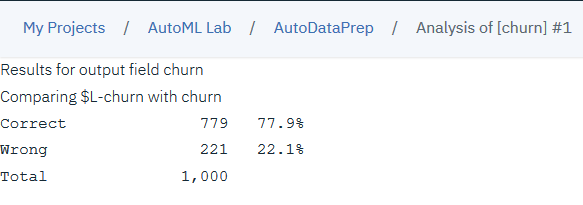
1. Agregue los nodos Auto Data **Prep** (la pestaña **Field Operations** ) y Logistic (la pestaña **Modeling)** al nodo **Type.**



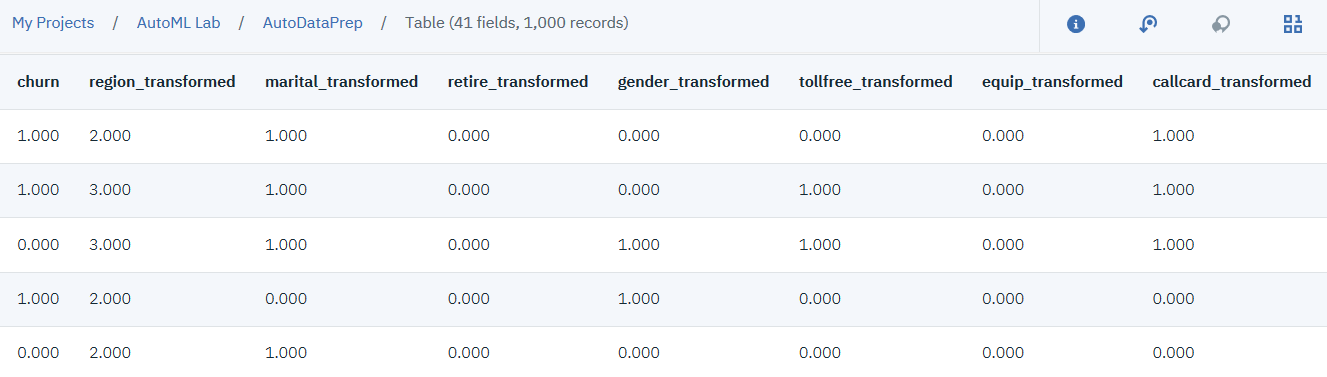
1. Haga clic con el botón derecho en nodo **Logístico** (conectado a **Auto Data** **Prep**) y seleccione **Ejecutar**.
2. Una vez creado el modelo, conecte los nodos **Tabla** y **Análisis** al modelo.



1. Ejecute el nodo **Análisis** que acaba de agregar. Observe que la precisión ha aumentado del 10% al 78%.



1. Ejecute el nodo **Tabla** y revise los campos que se han transformado.



Las transformaciones realizadas por el nodo **Auto Data Prep** se describen en la guía **Algoritmos:** <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.2/en/AlgorithmsGuide.pdf>

En un nivel alto, **Auto Data Prep** utiliza varias técnicas de evaluación, transformación y selección de características,

* colección de estadísticas univariadas
* measurement level rcasting
* missing value handling
* outlier identificación y handling
* continuous predictor transformations
* target handling
* reordering ctegories
* bivariate statistics collection
* supervised binning
* elección y construction
* importancia dela característica.

Como se mencionó anteriormente, los científicos de datos pueden realizar varias acciones después de completar la preparación automática de datos. Pueden seguir trabajando en Modeler o exportar los datos y crear modelos en código abierto.

Ha terminado el AutoML en el laboratorio de Watson Studio.