

# BUSINESS & TECH SCHOOL UNIVERSIDAD EUROPEA CON IBM

**Where business  
meets technology.**

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>

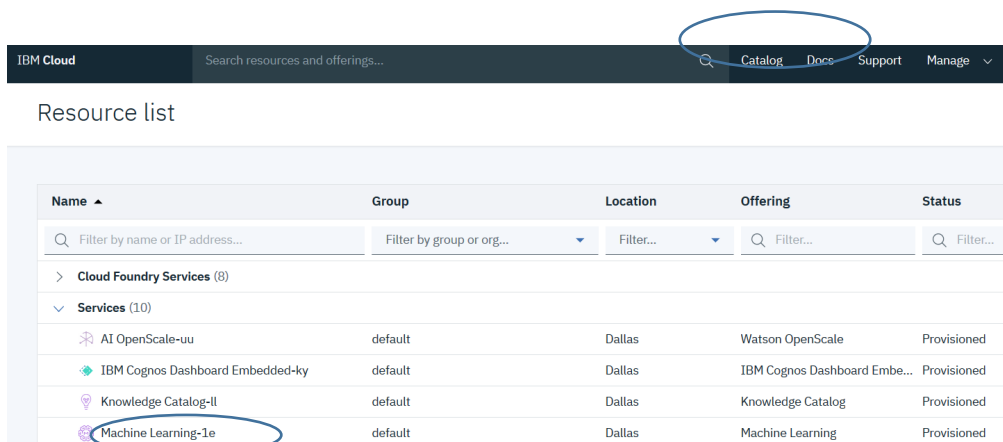
## Software, acceso y archivos necesarios

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.



The screenshot shows the IBM Cloud 'Catalog' page. At the top, there's a search bar and navigation links for 'Catalog', 'Docs', 'Support', and 'Manage'. Below the search bar, the 'Resource list' is displayed. It includes a table with columns: Name, Group, Location, Offering, and Status. Under the 'Services (10)' section, the 'Machine Learning-1e' service is listed and circled in blue. The table also shows other services like 'AI OpenScale-uu', 'IBM Cognos Dashboard Embedded-ky', and 'Knowledge Catalog-ll'.

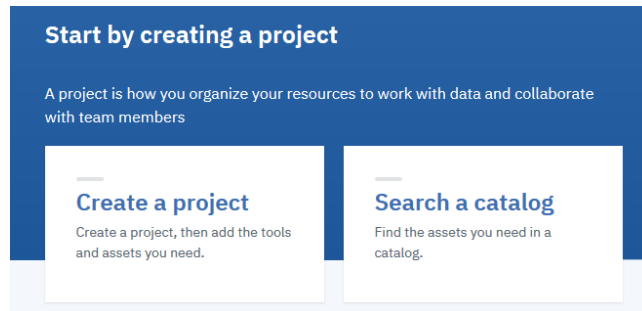
Name	Group	Location	Offering	Status
AI OpenScale-uu	default	Dallas	Watson OpenScale	Provisioned
IBM Cognos Dashboard Embedded-ky	default	Dallas	IBM Cognos Dashboard Embe...	Provisioned
Knowledge Catalog-ll	default	Dallas	Knowledge Catalog	Provisioned
Machine Learning-1e	default	Dallas	Machine Learning	Provisioned

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

## Parte 1: Generación de pipelines AutoML

**Watson Studio** proporciona varias capacidades de *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**



3. Seleccione **Crear un proyecto vacío**.



### Create an empty project

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

**NEW** AutoAI experiment tool: Fully automated approach to building a classification or re...

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

New project

---

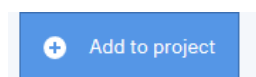
Define project details

**Name**

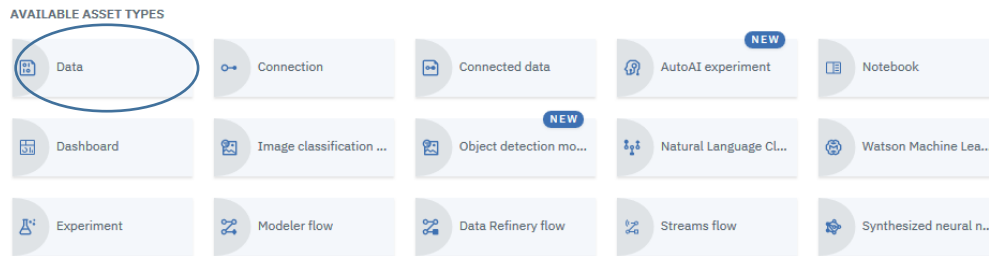
AutoML Lab

**Description**

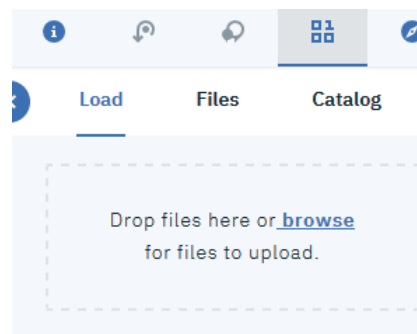
5. A continuación, cargaremos los activos de datos que se usarán para el modelado. Los archivos .csv se encuentran en el directorio **/data** del archivo que descargó de **GitHub**.
  - a. En **Project** view seleccione Agregar al proyecto -> **Datos**.



## Choose asset type



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



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

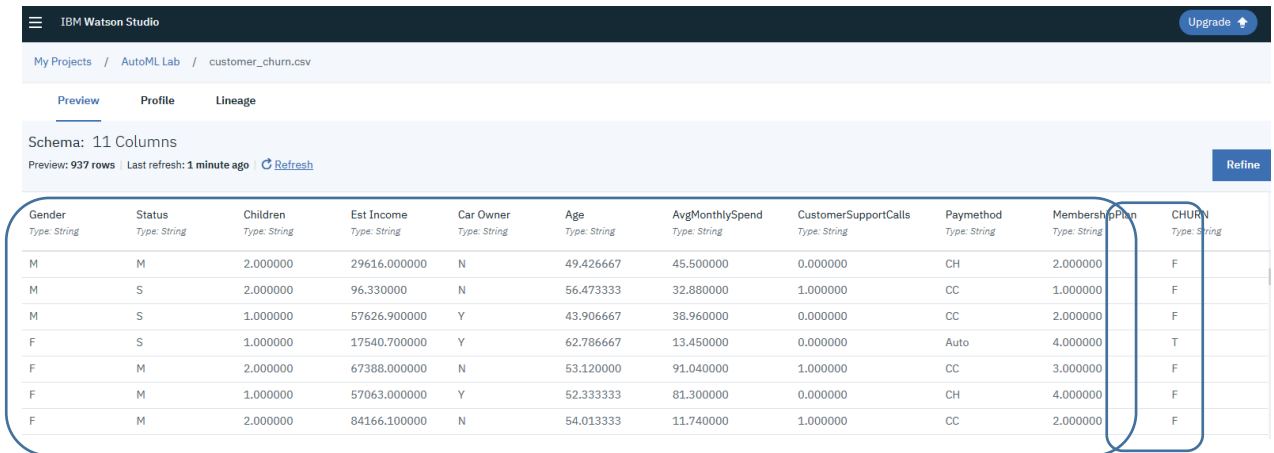
Data assets		
0 asset selected.		
<input type="checkbox"/>	NAME	TYPE
<input type="checkbox"/>	CSV <a href="#">customer_data.csv</a>	Data Asset
<input type="checkbox"/>	CSV <a href="#">telco.csv</a>	Data Asset
<input type="checkbox"/>	CSV <a href="#">customer_churn.csv</a>	Data Asset

6. 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.

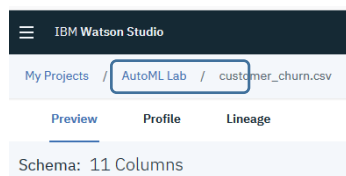
Con AutoAI, crearemos un modelo que predecirá la renovación para nuevos clientes. Un departamento de marketing puede utilizar esta información para prevenir de forma proactiva el abandono.

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*.

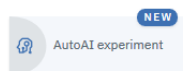


Gender	Status	Children	Est Income	Car Owner	Age	AvgMonthlySpend	CustomerSupportCalls	Paymethod	MembershipPlan	CHURN
M	M	2.000000	29616.000000	N	49.426667	45.500000	0.000000	CH	2.000000	F
M	S	2.000000	96.330000	N	56.473333	32.880000	1.000000	CC	1.000000	F
M	S	1.000000	57626.900000	Y	43.906667	38.960000	0.000000	CC	2.000000	F
F	S	1.000000	17540.700000	Y	62.786667	13.450000	0.000000	Auto	4.000000	T
F	M	2.000000	67388.000000	N	53.120000	91.040000	1.000000	CC	3.000000	F
F	M	1.000000	57063.000000	Y	52.333333	81.300000	0.000000	CH	4.000000	F
F	M	2.000000	84166.100000	N	54.013333	11.740000	1.000000	CC	2.000000	F

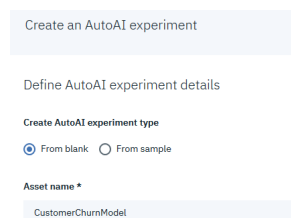
7. Haga clic en **Auto ML Lab** para volver a la vista **Proyecto** .



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



9. Asigne al activo el nombre *CustomerChurnModel*.



Create an AutoAI experiment

Define AutoAI experiment details

Create AutoAI experiment type

☒ From blank ☐ From sample

Asset name \*

CustomerChurnModel

10. A continuación, tenemos que agregar el **archivo WML** a nuestro proyecto.

- a. Haga clic en **Asociar...** enlace.

#### Associated services

##### Machine Learning Service

No Machine Learning service instances associated with your project.

[Associate a Machine Learning service instance](#) with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

[Reload](#)


##### Compute configuration \*

8 vCPU and 32 GB RAM

This compute configuration consumes 20 capacity units per hour. [Learn more](#) about capacity unit hours and Watson Machine Learning pricing plans.

- b. Seleccione el **servicio WML**

Existing Service Instance


Machine Learning-1e 

[Select](#) [Cancel](#)

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

#### Associated services

##### Watson Machine Learning Service Instance \*

Machine Learning-1e 

##### Compute configuration \*

8 vCPU and 32 GB RAM

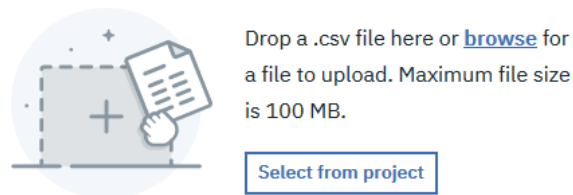
This compute configuration consumes 20 capacity units per hour. [Learn more](#) about capacity unit hours and Watson Machine Learning pricing plans.

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

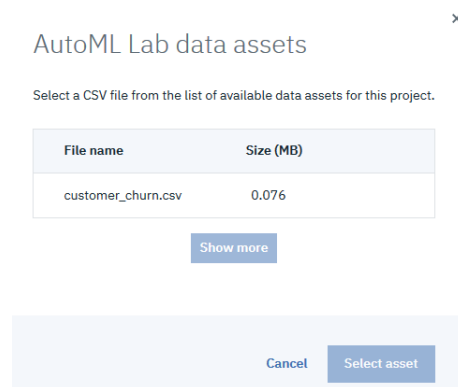




11. En la pantalla Agregar datos de **entrenamiento**, haga clic en **Seleccionar del proyecto**



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



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

### Select column to predict

DATA SOURCE: customer\_churn.csv

Column name	Type
Age	Integer
AvgMonthlySpend	Decimal
CustomerSupportCalls	Integer
Paymethod	String
MembershipPlan	Integer
CHURN	String

Selected prediction

PREDICTION TYPE

**Binary Classification**

POSITIVE CLASS

**T**

OPTIMIZED METRIC

**ROC AUC**

[Experiment settings](#)
[Run experiment](#)

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.

PREDICTION TYPE

Regression 

OPTIMIZED METRIC

RMSE 

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


Experiment settings

Data source

Prediction

General

Data source settings


PREDICTION COLUMN 

CHURN

COLUMN DATA TYPE

String

DATA SOURCE

customer\_churn.csv 


Subsample

For a large data set, use a subset of data to train the experiment. This speeds up results but may affect accuracy.

☐ Subsample rows

Training data split

You can optionally adjust the percentage of your data source to use for creating, optimizing, and validating pipelines. Only recommended for large data sets to avoid decreasing the quality of the pipelines.

85%  95%

Training data split: 90% — 3 folds | Holdout data split: 10%

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.

15. 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**.

Experiment settings
Data source
**Prediction**
General

### Prediction settings

PREDICTION COLUMN ⓘ  
CHURN

COLUMN DATA TYPE  
String

DATA SOURCE  
customer\_churn.csv ⓘ

**Prediction type**  
Change the prediction type based on data in the prediction column. Changing the type changes other prediction settings.

**Binary classification** ⓘ  
Classify data into categories. Choose this if your prediction column contains two distinct categories.

**Multiclass classification**  
Classify data into categories. Choose this if your prediction column contains multiple distinct categories.

**Regression**  
Predict values from a continuous set of values. Choose this if your prediction column contains a large number of values.

**Positive class**  
Specify the value in your prediction column to measure performance in to a confusion matrix.

T

**Optimized metric**  
Choose the metric to optimize for the experiment.

ROC AUC (Recommended) ⓘ
Accuracy
Average Precision
Precision
Recall
F1
Log Loss

**Algorithms to include**  
Select which of the following algorithms is to be considered when the experiment is run. The list of algorithms are based on the selected prediction type.

Search by algorithm

Included algorithms 7/7

16. 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.

Experiment settings  
Data source  
Prediction  
**General**

### General settings

**Experiment details**  
Review additional details about your experiment. These cannot be changed.

Branching factor (4 pipelines per branch):	1
Initial model tuning iterations:	25
Feature engineering iterations:	60
Final model tuning iterations:	50

**Watson Machine Learning service**

Watson Machine Learning - Demo Environment

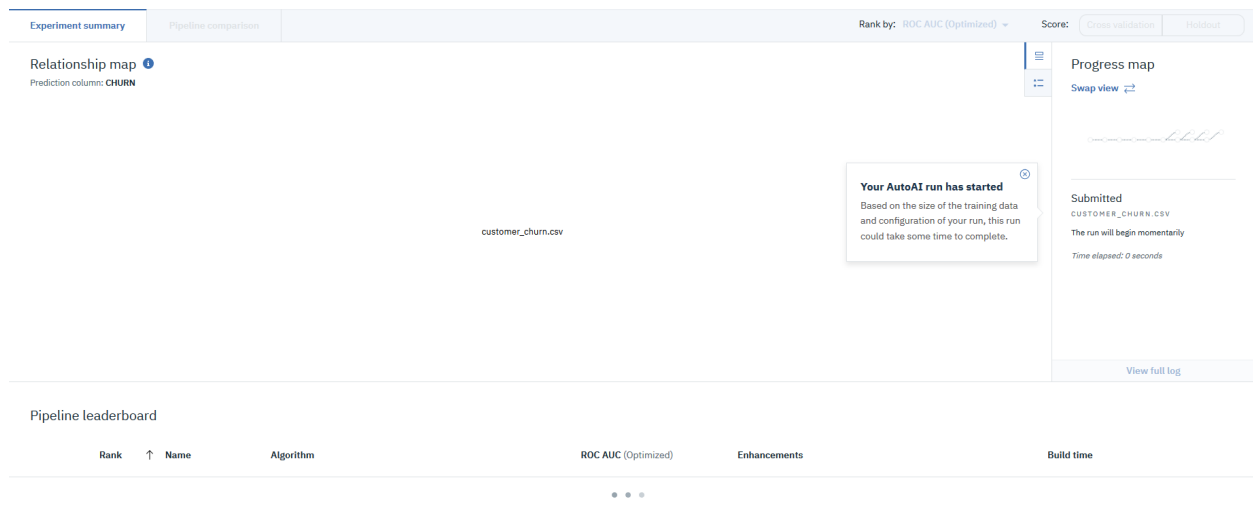
**Compute configuration**  
The selected runtime has 8 vCPU and 32 GB RAM and consumes 20 capacity units per hour.  
[Learn more](#) about capacity unit hours and Watson Machine Learning pricing plans.

8 vCPU and 32 GB RAM

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

PREDICTION TYPE	POSITIVE CLASS	OPTIMIZED METRIC
<b>Binary Classification</b> ⓘ	<b>T</b>	<b>ROC AUC</b> ⓘ
Experiment settings ⚙️		Run experiment ▶️

## 19. 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.

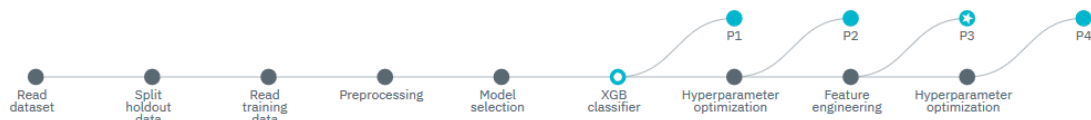
El siguiente estado que se muestra, una vez que el trabajo se ha configurado es leer los datos

Reading data source  
CUSTOMER\_CHURN.CSV  
Downloading source data  
Time elapsed: 22 seconds

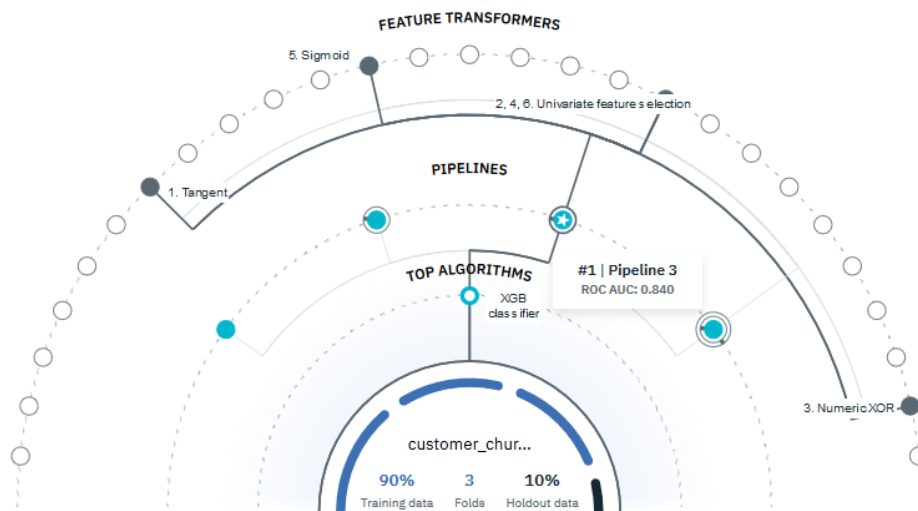
Durante el procesamiento, es posible cambiar entre dos tipos de salida mediante el botón Intercambiar vista:

**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



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

Pipeline leaderboard

	Rank	↑	Name	Algorithm	ROC AUC (Optimized)	Enhancements	Build time
>	★ 1		<a href="#">Pipeline 3</a>	XGB classifier	0.840	HPO-1 FE	00:06:15
>	2		<a href="#">Pipeline 4</a>	XGB classifier	0.840	HPO-1 FE HPO-2	00:02:11
>	3		<a href="#">Pipeline 2</a>	XGB classifier	0.830	HPO-1	00:00:46
>	4		<a href="#">Pipeline 1</a>	XGB classifier	0.829	None	00:00:02

21. 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**:

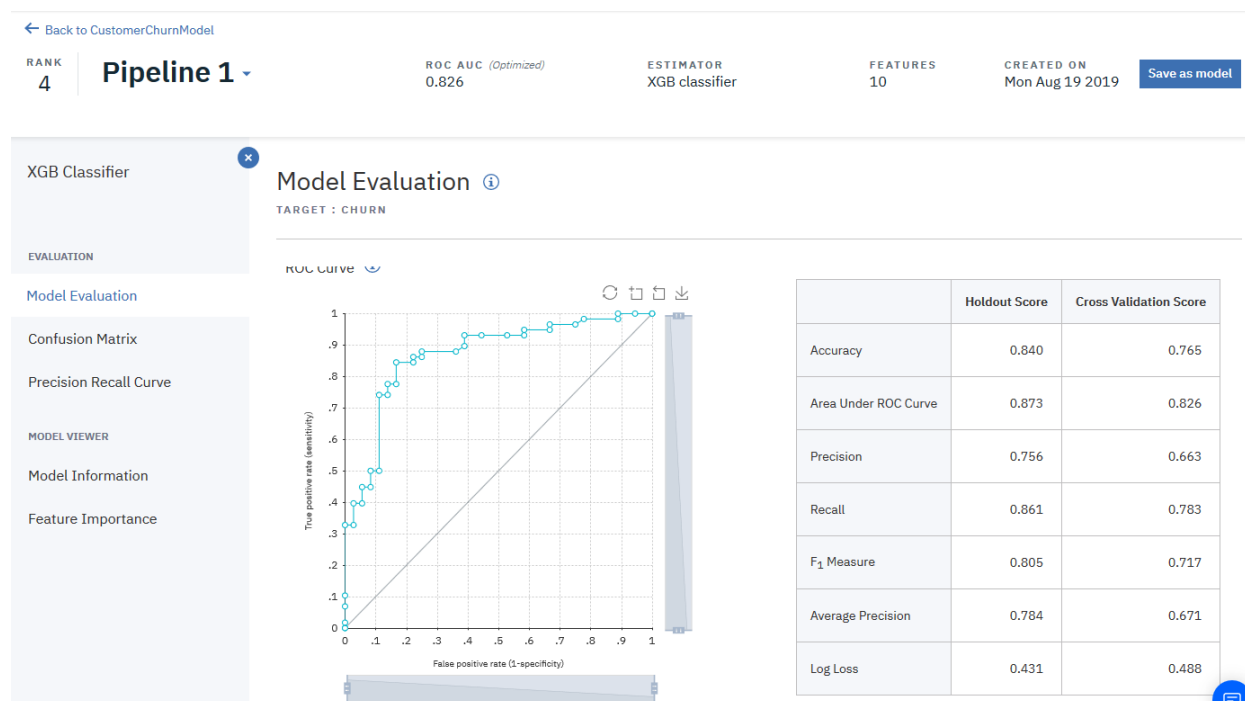
*.90-1 á excelente (A)*

*.80-.90 á bueno (B)*

*.70-.80 á feria (C)*

*.60-.70 á pobres (D)*

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



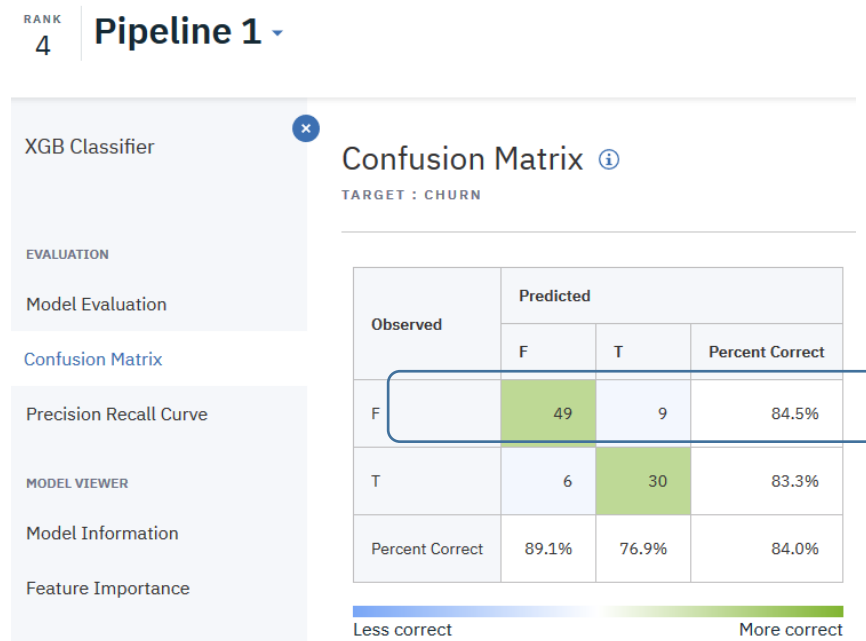




22. Haga clic en **Matriz de confusión** a la izquierda.

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

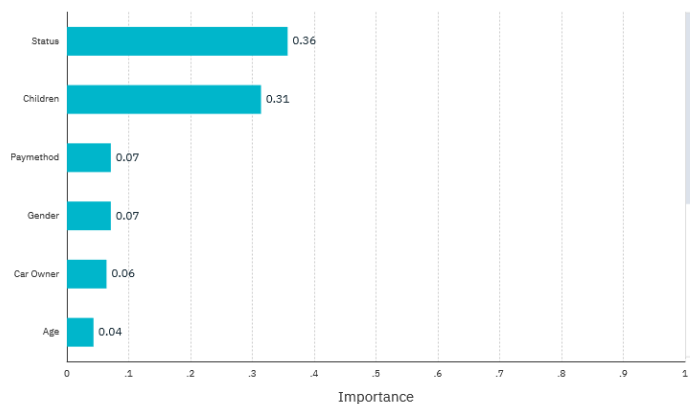
Por ejemplo, el modelo predijo 49 de F (False – cliente no churn) registros como F (predicción correcta). Al mismo tiempo, predijo 9 registros as T (el cliente va a abandonar), pero el valor real era F. En general, para la predicción False la precisión es del 84%.



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

## Feature Importance ⓘ

TARGET : CHURN



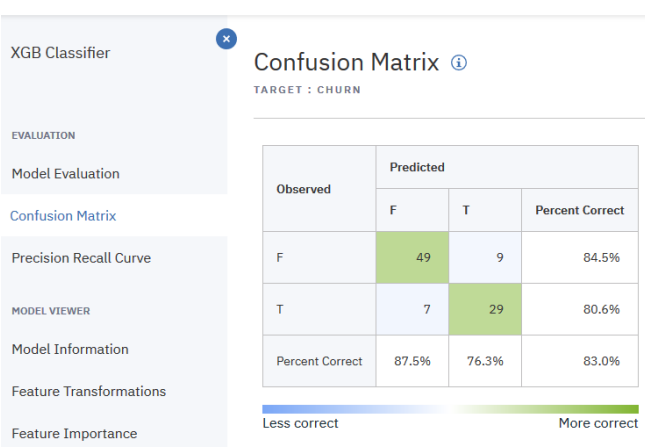
24. A continuación, vuelva a la vista **pipelines** y haga clic en pipeline **4**.

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.

RANK 2 Pipeline 4 ▾



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

RANK 2 Pipeline 4

XGB Classifier

EVALUATION

Model Evaluation

Confusion Matrix

Precision Recall Curve

MODEL VIEWER

Model Information

Feature Transformations

Feature Importance

Feature Transformations ⓘ

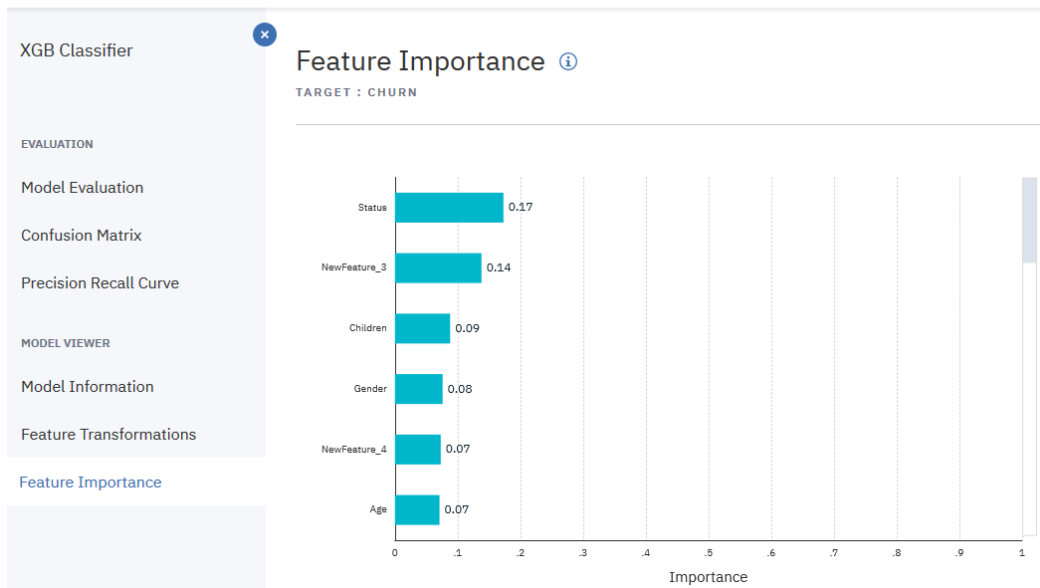
TARGET : CHURN

New Feature	Original Feature	Transformation
NewFeature_3	Est Income,Gender	$\text{rxor}(\text{Est Income}, \text{Gender})$
NewFeature_4	Est Income,Age	$\text{rxor}(\text{Est Income}, \text{Age})$
NewFeature_0	Status,Children	$\text{rxor}(\text{Status}, \text{Children})$
NewFeature_2	Paymethod,Children	$\text{rxor}(\text{Children}, \text{Paymethod})$
NewFeature_6	Age,AvgMonthlySpend	$\text{rxor}(\text{Age}, \text{AvgMonthlySpend})$
NewFeature_11	Est Income,Age	$\text{sigmoid}(\text{rxor}(\text{Age}, \text{Est Income}))$
NewFeature_1	Status,Children	$\text{rxor}(\text{Children}, \text{Status})$

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.

RANK  
2 Pipeline 4 ▾



25. 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**.

Pipeline leaderboard

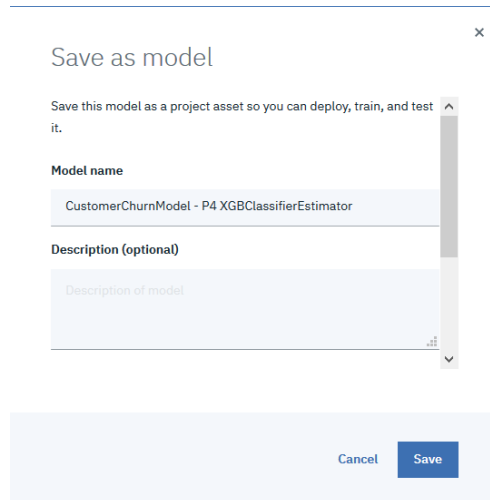
[Compare pipelines](#)

Ranking based on:

ROC AUC

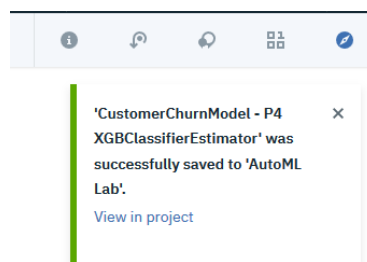
RANK	PIPELINE	ROC AUC	ESTIMATOR	ENHANCEMENTS
> 1	<a href="#">Pipeline 4</a>	0.830	XGB classifier	HPO FE HPO + FE <a href="#">Save as model</a>

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



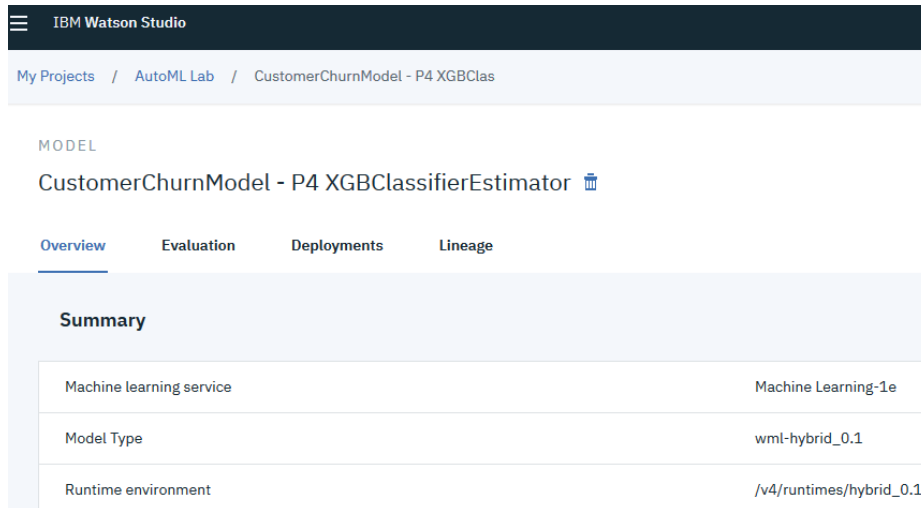
A dialog box titled "Save as model" with a close button (X) in the top right corner. Below the title, it says "Save this model as a project asset so you can deploy, train, and test it." There are two input fields: "Model name" with the text "CustomerChurnModel - P4 XGBClassifierEstimator" and "Description (optional)" with the placeholder text "Description of model". At the bottom, there are "Cancel" and "Save" buttons.

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



26. A continuación, configuraremos la implementación.

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



IBM Watson Studio

My Projects / AutoML Lab / CustomerChurnModel - P4 XGBClas

MODEL

CustomerChurnModel - P4 XGBClassifierEstimator

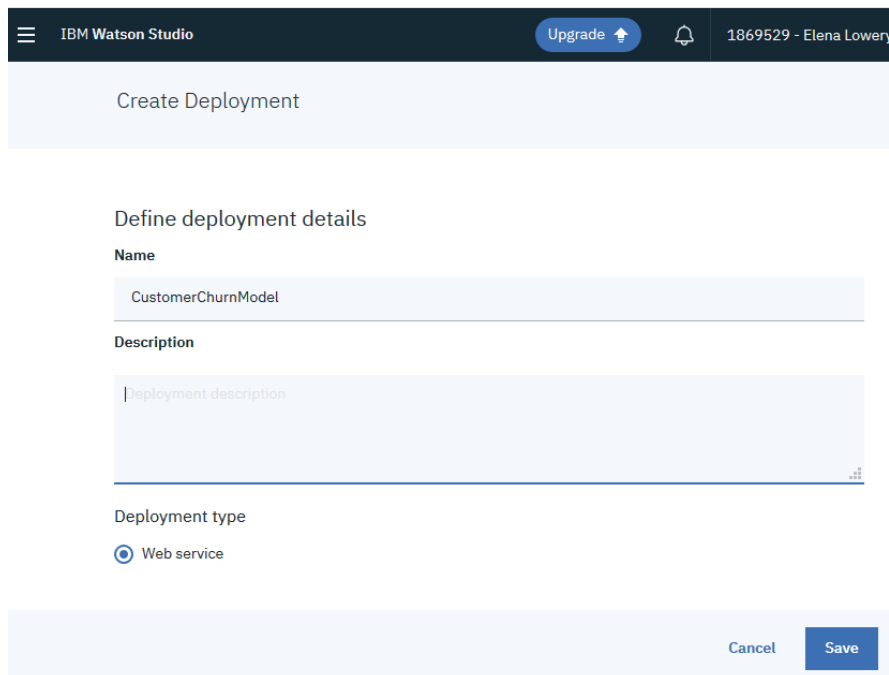
Overview Evaluation Deployments Lineage

**Summary**

Machine learning service	Machine Learning-1e
Model Type	wml-hybrid_0.1
Runtime environment	/v4/runtimes/hybrid_0.1

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

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



IBM Watson Studio Upgrade 1869529 - Elena Lowery

Create Deployment

Define deployment details

**Name**

CustomerChurnModel

**Description**

Deployment description


**Deployment type**

☒ Web service

Cancel Save

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

MODEL

CustomerChurnModel - P4 XGBClassifierEstimator 

Overview

Evaluation

Deployments

Lineage

Add Deployment 

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
CustomerChurnModel	initializing	Web Service	



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

**CustomerChurnModel**

Overview Implementation **Test**

Enter input data

Gender

Status

Children

Est Income

Predict

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 servicios con una confianza del 73 por ciento.

Overview Implementation **Test**

Enter input data

7.3

CustomerSupportCalls

0

Paymethod

CH

MembershipPlan

2

Predict

```
{
  "predictions": [
    {
      "fields": [
        "prediction",
        "probability"
      ],
      "values": [
        "F",
        [
          0.7337597608566284,
          0.2662402093410492
        ]
      ]
    }
  ]
}
```

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**.

## CustomerChurnModel

Overview **Implementation** Test

### Implementation

Scoring End-point	<a href="https://us-south.ml.cloud.ibm.com/v4/deployments/e6a238c3-661f-4a6c-995b-bff1b5dcb6d6/predictions">https://us-south.ml.cloud.ibm.com/v4/deployments/e6a238c3-661f-4a6c-995b-bff1b5dcb6d6/predictions</a>
Authorization: Bearer <token>	Review the <a href="#">WML authentication</a> documentation for details about generating IAM tokens.
ML-Instance-ID	The "ML-Instance-ID" HTTP header must be populated with the WML instance id, which can be obtained as <a href="#">described here</a> .
Content-type: application/json	Required if the request body is sent in JSON format.

### Code Snippets

cURL Java JavaScript **Python** Scala

```
import urllib3, requests, json

# NOTE: generate iam_token and retrieve ml_instance_id based on provided documentation
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + iam_token, 'ML-Instance-ID': ml_instance_id}
```

## 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 apaso):* 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:  
<http://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.2/en/AlgorithmGuide.pdf>

SPSS Modeler contiene varias características de AutoAI:

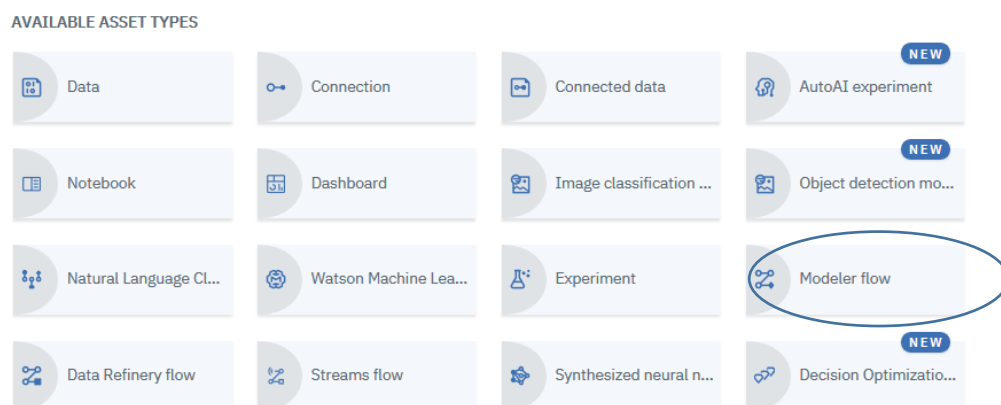
- Modelado automático para modelos de clasificación, regresión y agrupación en clústeres
- Preparación automática de datos
- Auditoría de datos
- Selección automática de funciones
- Características automáticas en algoritmos individuales (por ejemplo, regresión).

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

### Modelado automático

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

#### Choose asset type



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

### New modeler flow

**New**
From File
From Example

**Name\***

AutoCustomerChurn

**Description**

| Type description here.

**Select flow type**

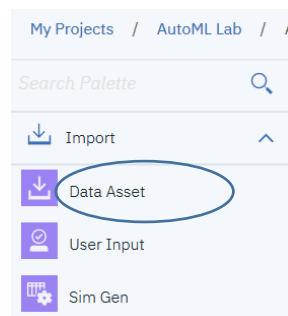
☒ Modeler Flow
 ☐ Neural Network Modeler BETA

**Runtime**

☒ IBM SPSS Modeler
 ☐ Spark BETA

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

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



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

Data Asset

DATA

Change data asset

Source location

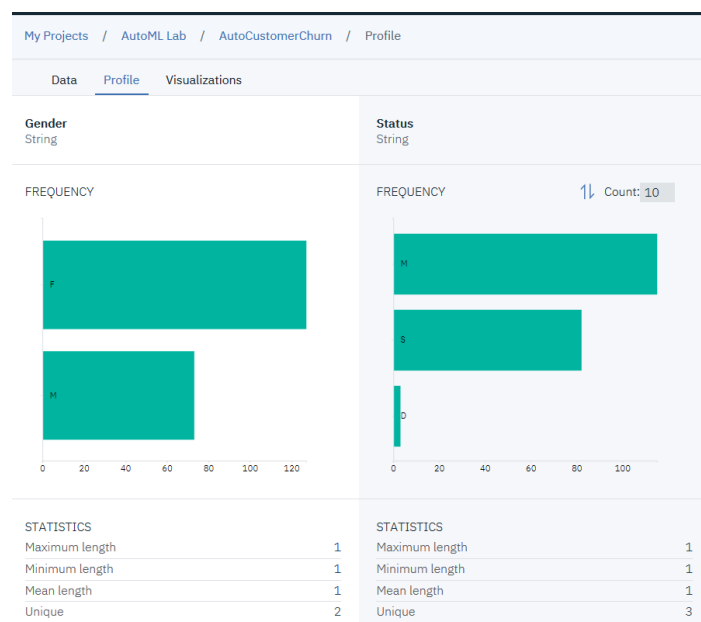
ANNOTATIONS

5. Expanda **Data Assets** y seleccione `customer_churn.csv`. Haga clic en **Aceptar**.

My Projects / AutoML Lab / AutoCustomerChurn	
<b>AutoML Lab</b>	<b>Data assets</b>
Assets (2)	Data assets (1)
Connections >	customer_churn.csv
Data assets >	

6. 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.

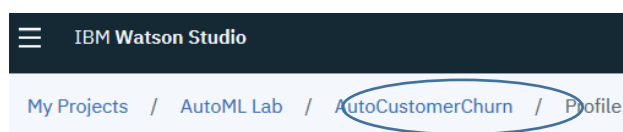
*Nota: si recibe el error `ElementLocked`, haga clic en **Aceptar**.*



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.

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



8. 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.

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

Type

---

SETTINGS

---

Default Mode ⓘ

☒ Read metadata ☐ Pass (do not scan)

> Type Operations

Read Values

Clear All Values

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.*)

Search in column Field

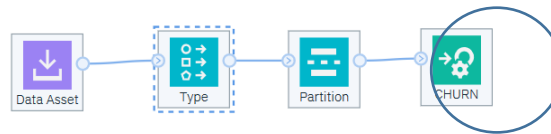
<input type="checkbox"/>	Field	Measure	Role	Value mode	Values
<input type="checkbox"/>	Gender	Flag	Input	Specify	F, M
<input type="checkbox"/>	Status	Nominal	Input	Specify	D, M, S
<input type="checkbox"/>	Children	Continuous	Input	Specify	0.0, 2.0

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

<input type="checkbox"/>	Field	Measure	Role	Value mode	Values
<input type="checkbox"/>	Est Income	Continuous	Input	Specify	96.33, 120000.0
<input type="checkbox"/>	Car Owner	Flag	Input	Specify	N, Y
<input type="checkbox"/>	Age	Continuous	Input	Specify	12.326667, 77.0
<input type="checkbox"/>	AvgMonthlyS...	Continuous	Input	Specify	0.79, 332.46
<input type="checkbox"/>	CustomerSu...	Continuous	Input	Specify	0.0, 4.0
<input type="checkbox"/>	Paymethod	Nominal	Input	Specify	Auto, CC, CH
<input type="checkbox"/>	Membership...	Continuous	Input	Specify	1.0, 4.0
<input type="checkbox"/>	CHURN	Flag	Target	Specify	F, T

Haga clic en **Guardar**.

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



10. 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**.

**Training Partition(%)** ⓘ

90

---

**Testing Partition(%)** ⓘ

10



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

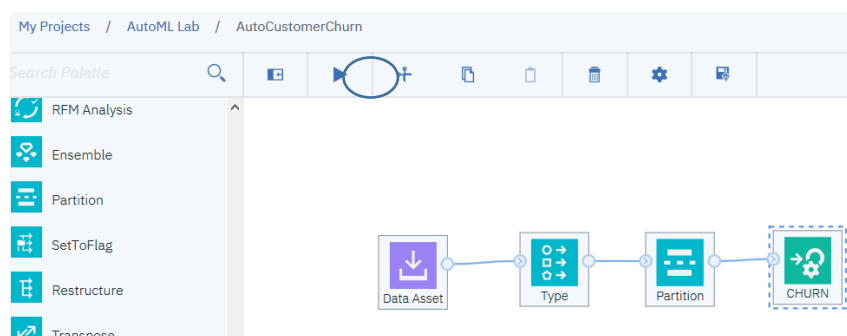
Un científico de datos experimentado puede modificar varias opciones en la configuración del nodo.

CHURN	
FIELDS	▼
BUILD OPTIONS	▼
CRITERIA FOR FLAG TARGETS	▼
EXPERT	▼
DISCARD	▼
ENSEMBLE	▼
ANNOTATIONS	▼

Usaremos la configuración predeterminada para el resto de los valores.

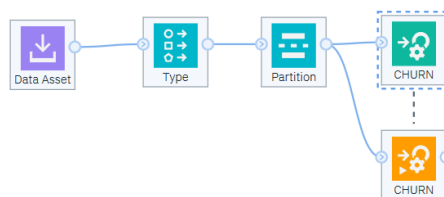
Haga clic en **Cancelar** para volver al lienzo.

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



13. 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.*



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

Modeler mostró los 3 modelos principales.

## Auto Classifier - Models ⓘ

TARGET : CHURN

USE	ESTIMATOR	ACCURACY	BUILD TIME (MINS)	NO. FIELDS USED	ACTIONS
<input checked="" type="checkbox"/>	<a href="#">CHAID</a>	79.348	< 1	7	
<input checked="" type="checkbox"/>	<a href="#">MLP Neural Network</a>	75.000	< 1	10	
<input checked="" type="checkbox"/>	XGBoost Tree 1	73.913	< 1	10	

15. 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**.

CHAID Tree Model ⓘ

MODEL VIEWER

Model Information

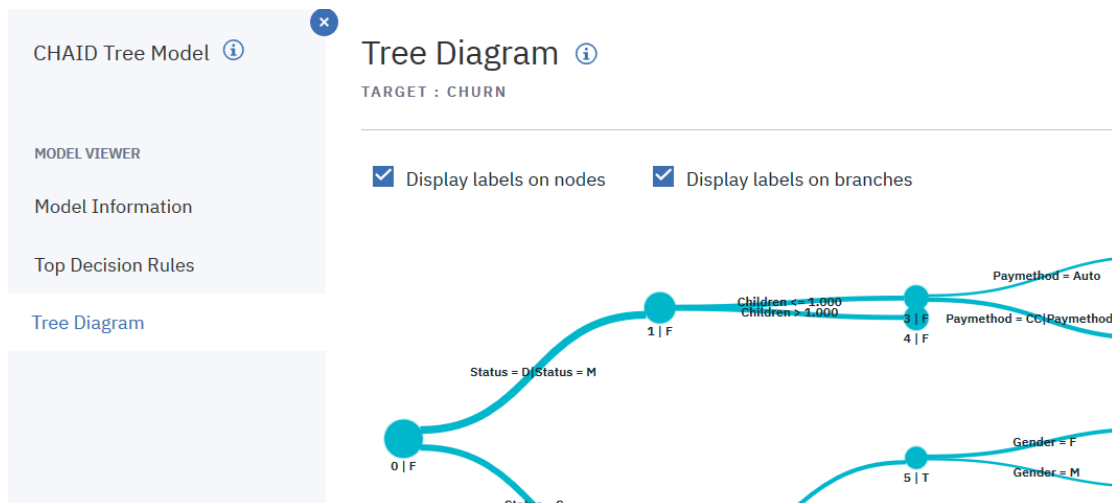
Top Decision Rules

Tree Diagram

Top Decision Rules ⓘ

TARGET : CHURN

Rule ID	Rule	Mode category	Record count	Record percentage	Rule confidence
4	((Status = D or Status = M)) and ((Children > 1.0))	F	240	28.402	93.750



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

### Auto Classifier - Models

TARGET : CHURN

USE	ESTIMATOR	ACCURACY	BUILD TIME (MINS)	NO. FIELDS USED	ACTIONS
<input checked="" type="checkbox"/>	<a href="#">CHAID</a>	79.348	< 1	7	
<input checked="" type="checkbox"/>	<a href="#">MLP Neural Network</a>	75.000	< 1	10	
<input checked="" type="checkbox"/>	XGBoost Tree 1	73.913	< 1	10	

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).

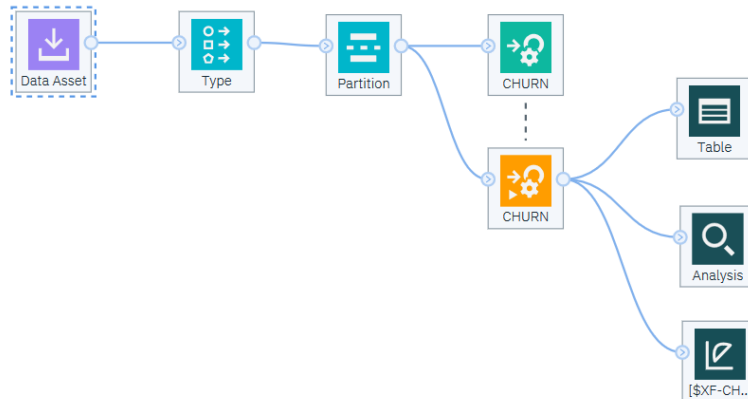
Por último, un científico de datos puede utilizar el modelado automático como un paso rápido para determinar qué modelo se debe utilizar. Más adelante pueden crear un modelo individual para el rendimiento superior (todos los algoritmos que se usan en nodos automáticos están disponibles como algoritmos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

Copyright IBM Corporation 2013-2018. All rights reserved.

16. 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.

17. 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.

My Projects / AutoML Lab / AutoCustomerChurn / Table (14 fields, 937 records)													
Gender	Status	Children	Est Income	Car Owner	Age	AvgMonthlySpend	CustomerSupportCalls	Paymethod	MembershipPlan	CHURN	Partition	\$XF-CHURN	\$XFC-CHURN
M	M	2.000	29616.000	N	49.427	45.500	0.000	CH	2.000	F	1_Training	F	0.949
M	S	2.000	96.330	N	56.473	32.880	1.000	CC	1.000	F	1_Training	F	0.821
M	S	1.000	57626.900	Y	43.907	38.960	0.000	CC	2.000	F	1_Training	F	0.933

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

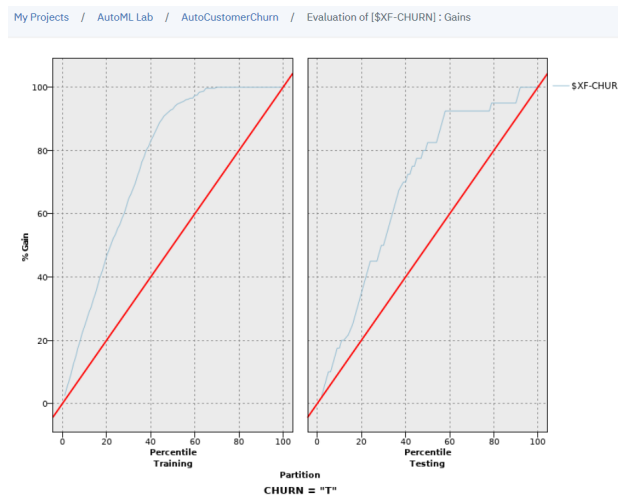


My Projects / AutoML Lab / AutoCustomerChurn / Analysis of [CHURN]

Results for output field CHURN  
Comparing \$XF-CHURN with CHURN

'Partition'	1_Training	2_Testing
Correct	718 84.97%	73 79.35%
Wrong	127 15.03%	19 20.65%
Total	845	92

19. 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*.

- Si se le pide que cambie el **Data Asset**, cargue el archivo *customer\_data.csv* desde el directorio */data*.

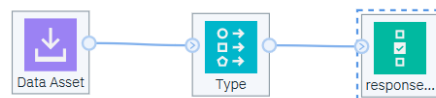
Si no ha visto el mensaje, haga doble clic en **Data Asset**, haga click en **Cambiar** Data Asset seleccione *customer\_data.csv* en el proyecto.

My Projects / TestAILab / FeatureSelection	
TestAILab	Data assets
Assets (2)	Data assets (3)
Connections	> customer_churn.csv
Data assets	> customer_data.csv
	telco.csv

- 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.

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



- Haga clic con el botón derecho en el nodo **Selección** de características y seleccione **Ejecutar**.
- 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.

response\_01

MODEL

<input type="checkbox"/>	Rank ^	Field ^	Measurement ^	Importance ^	Value ^
<input checked="" type="checkbox"/>	1	ed	Continuous	Important	1.000
<input checked="" type="checkbox"/>	2	edcat	Continuous	Important	1.000
<input checked="" type="checkbox"/>	3	ownpc	Continuous	Important	1.000
<input checked="" type="checkbox"/>	4	internet	Continuous	Important	1.000
<input checked="" type="checkbox"/>	5	equip	Continuous	Important	1.000

Selected fields: 31 Total fields available: 130  
☒ > 0.95 ☐ <= 0.95 ☐ < 0.9  
 7 Screened Fields

<input type="checkbox"/>	Field ^	Measurement ^	Reason ^
<input type="checkbox"/>	Inequipmon	Continuous	Coefficient of variation below threshold
<input type="checkbox"/>	Inwiremon	Continuous	Too many missing values
<input type="checkbox"/>	Inwireten	Continuous	Too many missing values
<input type="checkbox"/>	owntv	Nominal	Single category too large
<input type="checkbox"/>	ownvcr	Nominal	Single category too large

- 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

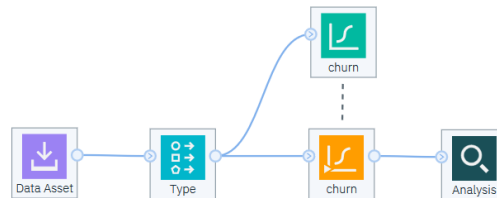
La preparación automática de datos en Modeler se puede utilizar como una tarea independiente o en combinación con la selección y el modelado de características.

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

- Vaya al proyecto y seleccione Agregar flujo **de modelador**.
- Seleccione **Desde archivo**. Vaya al directorio `/flows` de los archivos de laboratorio descomprimidos e importe la secuencia *AutoDataPrep*.
- 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.

- 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**.



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

La precisión es pobre.

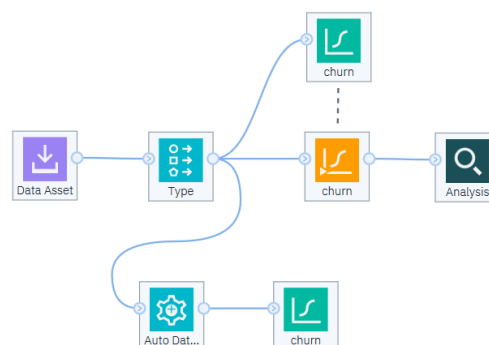
My Projects / AutoML Lab / AutoDataPrep / Analysis of [churn]

Results for output field churn

Comparing \$L-churn with churn

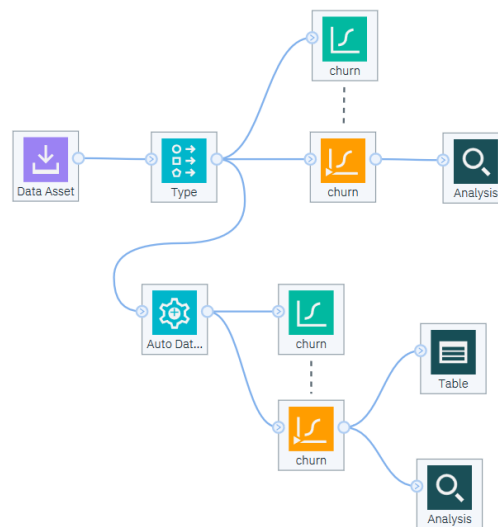
Correct	106	10.6%
Wrong	894	89.4%
Total	1,000	

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



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





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

My Projects / AutoML Lab / AutoDataPrep / Analysis of [churn] #1			
Results for output field churn			
Comparing \$L-churn with churn			
Correct	779	77.9%	
Wrong	221	22.1%	
Total	1,000		

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

My Projects / AutoML Lab / AutoDataPrep / Table (41 fields, 1,000 records)							
churn	region_transformed	marital_transformed	retire_transformed	gender_transformed	tollfree_transformed	equip_transformed	callcard_transformed
1.000	2.000	1.000	0.000	0.000	0.000	0.000	1.000
1.000	3.000	1.000	0.000	0.000	1.000	0.000	1.000
0.000	3.000	1.000	0.000	1.000	1.000	0.000	1.000
1.000	2.000	0.000	0.000	1.000	0.000	0.000	0.000
0.000	2.000	1.000	0.000	0.000	0.000	0.000	0.000

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 recasting
- missing value handling
- outlier identificación y handling
- continuous predictor transformations
- target handling
- reordering categories
- bivariate statistics collection
- supervised binning
- selección y construcción
- importancia de la 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.