

# TAREA 4

Sophia Contreras

Yoksan Varela

Mauro Víquez

Por la naturaleza de la programación en la nube, el profesor nos recomendó hacer un trabajo investigativo. Por lo que se explica el funcionamiento de Google Cloud, AWS y además se da un ejemplo de ML con servicios en la nube de Microsoft Azure.

## GOOGLE CLOUD

### ¿Qué es Google Cloud?

Aunque su nombre lo sugiere, Google Cloud no es solamente almacenamiento en la nube; se trata de una infraestructura del tipo E2E (End-to-End), donde se pueden crear soluciones de completas para proyectos que responden a necesidades concretas, las cuales van a usar la infraestructura computacional que Google tiene a disposición. Literalmente, se rentan los recursos estructurales de Google para crear una solución para un proyecto.

Siendo más específicos, Google Cloud contiene muchas herramientas a disposición y el cliente decide, dentro de la arquitectura que quiera hacer para su proyecto, cuales de esas herramientas necesita y renta las mismas.

Existen muchas categorías para estas herramientas (y de diferentes complejidades), entre las cuales podemos mencionar (en este enlace se encuentra un mapa de todas las herramientas disponibles: [Google Cloud Developer Cheat Sheet \(googlecloudcheatsheet.withgoogle.com\)](https://googlecloudcheatsheet.withgoogle.com)):

- **Compute:** Contenedores y Máquinas Virtuales escalables.
- **Storage:** Almacenamiento para corto y largo plazo.
- **Database:** Base de datos relacionales o no-relacionales.
- **Data Analytics:** recolección, almacenaje, procesado y análisis de datos.
- **AI/ML:** Crear y usar soluciones de Machine Learning.

La siguiente imagen (tomada del enlace anterior) da una idea de la cantidad de herramientas disponibles:



ILUSTRACIÓN 1. HERRAMIENTAS DISPONIBLES EN GOOGLE CLOUD

Para mayor detalle, visitar el siguiente enlace que contiene toda la documentación oficial por parte de Google: [Google Cloud documentation | Documentation](https://cloud.google.com/documentation)

Con respecto al precio de hacer un proyecto con Google Cloud, este varía dependiendo de las herramientas que se necesiten para completar el proyecto. Se pueden realizar pagos mensuales o hacer un pago completo para un proyecto. En ambos casos, estamos hablando de miles de dólares, pero para hacerse de una idea más clara, se cuenta con una calculadora donde se pueden ir introduciendo herramientas y ver como varía el precio. El enlace a esta calcula es el siguiente: [Google Cloud Pricing Calculator](https://cloud.google.com/pricing-calculator)

## ¿En qué consiste Machine Learning en Google Cloud?

A continuación, se muestra un resumen del siguiente video en YouTube, creado por Google Cloud Tech: [Machine learning on Google Cloud \(youtube.com\)](https://www.youtube.com/watch?v=...).

Google pone a disposición herramientas que van desde soluciones completas y hechas por Google que pueden ser usadas directamente sin necesidad de tener conocimiento profundo en Machine Learning, así como las herramientas de infraestructura que permiten al cliente crear sus soluciones:

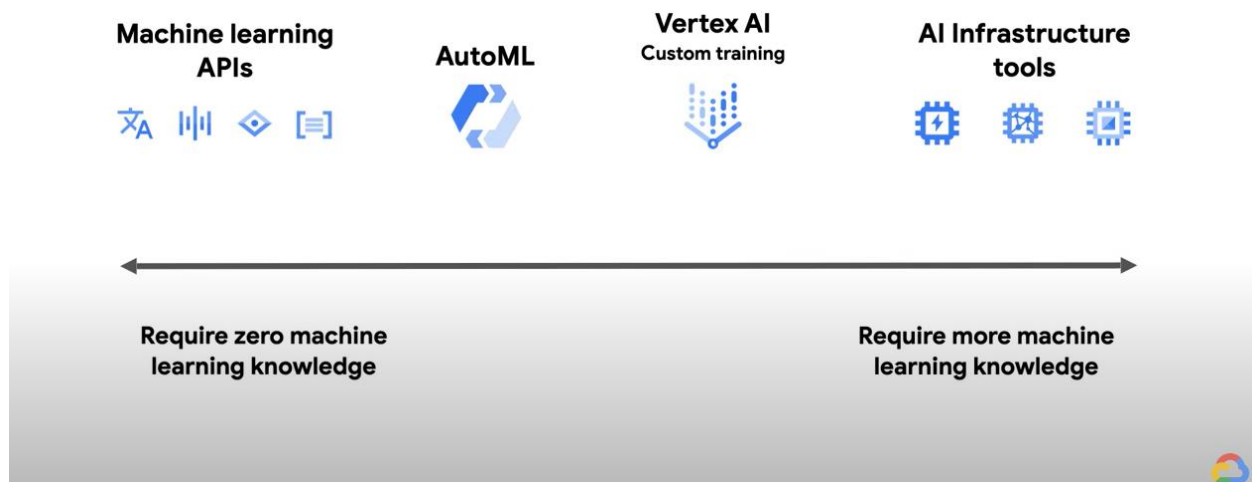


ILUSTRACIÓN 2. GENERALIDAD DE LAS HERRAMIENTAS DE ML DISPONIBLES EN GOOGLE CLOUD

## Machine Learning APIs

Son Interfaces para Programación Aplicada (API por sus siglas en inglés: Application Programming Interface) ya creadas por Google, con modelos de Machine Learning previamente entrenados por ellos y desarrollos en conjunto con otros clientes. Hay varios tipos de APIs: para procesamiento de imagen, video, traducción a otros idiomas, reconocimiento de voz, convertidos de texto hacia voz, entre otros.

En este caso, no se requiere conocimiento previo en Machine Learning dado que Google Cloud se encarga de la recolección de datos, la generación del modelo y el mantenimiento de este; por lo tanto, el cliente solo tiene que enfocarse en el análisis de los resultados.

Estos APIs pueden ser utilizados por diferentes lenguajes de programación para acoplarse a las necesidades del cliente. Algunos de estos lenguajes son: Python, Java, .NET, y muchos más.

## AutoML

Los datos de entrenamiento para el modelo son entregados por el cliente, y Google Cloud crea el modelo de Machine Learning con el requerimiento de predicción necesarios para el cliente. Igual que en el caso de la APIs, hay diferentes versiones de AutoML: para visión, reconocimiento natural de lenguaje, para datos tabulados, entre otros.

La opción para acceder a estas herramientas es a través de Vertex AI, una opción disponible en la consola de desarrollo, la cual es explicada a continuación.

## Vertex AI

Es una plataforma de manejo de modelos para Machine Learning, que permite desarrollar, lanzar y escalar estos modelos sobre la plataforma de Google Cloud. La cantidad de código necesaria para realizar esta función es menor a otros competidores, donde ellos estiman que se necesita 8% menos de líneas de código.

Vertex AI está completamente integrado con TensorFlow, PyTorch y Scikit Learn; y también se cuenta con soporte o integración de Notebooks, Google Cloud console y las librerías como APIs presentes en la plataforma.

Además de herramientas para desarrollo, Vertex AI cuenta con MLOps tools, necesarias para el mantenimiento y monitoreo de los modelos desarrollados sobre esta estructura.

## AI Infrastructure Tools

A este punto, el cliente quiere tener aún más control sobre la solución y el desarrollo del proceso como tal, por lo que Google Cloud provee herramientas que permite la creación completa y mantenimiento de los modelos de Machine Learning. Estas herramientas se pueden dividir en 3 grupos principales:

- **Deep Learning VM Images:** Frameworks pre-instalados de TensorFlow, PyTorch y sk-learn listo para ser utilizados.
- **Cloud GPUs:** Muy útiles para acelerar procesos que requieran poder computacional como Visualizaciones en 3D, entrenamiento de modelos pesados o desarrollo científico computarizado.
- **Cloud TPUs:** Para entrenamiento de modelos de Machine Learning.

# AMAZON WEB SERVICES (AWS)

## ¿Qué es Amazon Web Services (AWS)?

A pesar de su nombre, Amazon (2024) describe que Amazon Web Services (AWS) va mucho más allá que simplemente almacenamiento en la nube. AWS es una plataforma completa de servicios en la nube que abarca desde tecnologías básicas de infraestructura, como cómputo, almacenamiento y bases de datos, hasta áreas emergentes como inteligencia artificial, machine learning, análisis de datos, y el Internet de las cosas (IoT). Se destaca por ofrecer una extensa variedad de servicios y características dentro de cada uno de ellos. [Cloud Products \(amazon.com\)](https://aws.amazon.com/cloud-products/)

Dentro del amplio espectro de servicios, AWS permite a los usuarios construir soluciones completas para proyectos específicos, utilizando la infraestructura computacional proporcionada por Amazon. De manera similar a Google Cloud, AWS facilita la creación de soluciones de extremo a extremo (E2E), donde los recursos estructurales se alquilan para adaptarse a las necesidades del proyecto.

En términos más específicos, AWS ofrece una profunda funcionalidad en sus servicios. Por ejemplo, en el ámbito de bases de datos, brinda una variedad de opciones diseñadas para diferentes aplicaciones, permitiendo a los usuarios seleccionar las herramientas más adecuadas para obtener el mejor rendimiento y costo.

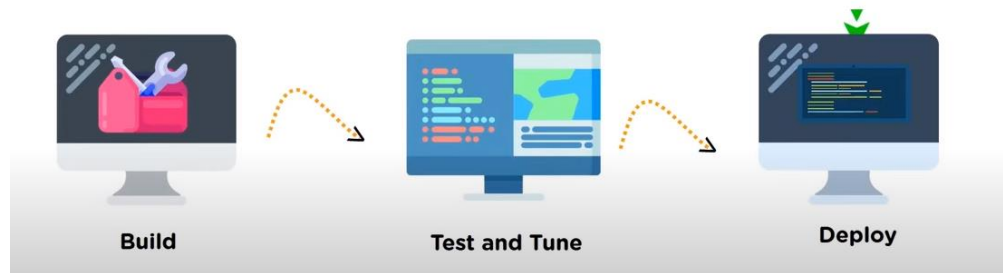
AWS también se destaca por tener la comunidad más grande y dinámica en la nube, con millones de clientes activos y miles de socios en todo el mundo. Empresas de todos los tamaños y sectores confían en AWS para ejecutar una amplia variedad de casos de uso.

Además, la plataforma de AWS está diseñada para ser flexible y segura, cumpliendo con los requisitos de seguridad para organizaciones de alta sensibilidad como el ámbito militar y bancario. Respaldo por un conjunto profundo de herramientas de seguridad en la nube y más de 300 servicios y funciones de seguridad, cumplen con 143 estándares de seguridad y certificaciones de cumplimiento. En resumen, AWS proporciona una experiencia incomparable en términos de amplitud y profundidad de servicios, seguridad, rendimiento y apoyo operativo, siendo una opción confiable para una amplia variedad de aplicaciones y proyectos en la nube.

En Amazon Web Services (AWS), Machine Learning (ML) es un conjunto de servicios y herramientas diseñadas para permitir a los desarrolladores y científicos de datos construir, entrenar e implementar modelos de machine learning de manera eficiente y escalable. Aquí hay algunas características clave de los servicios de Machine Learning en AWS [Getting Started with AWS Cloud Essentials \(amazon.com\)](https://aws.amazon.com/getting-started/machine-learning/):

- **Amazon SageMaker:**
  - **Entrenamiento de Modelos:** Amazon SageMaker facilita el entrenamiento de modelos de ML utilizando conjuntos de datos, con la posibilidad de escalar horizontalmente para mejorar el rendimiento.
  - **Despliegue de Modelos:** Permite implementar fácilmente modelos entrenados en entornos de producción para realizar inferencias en tiempo real.

Amazon SageMaker se puede resumir a través de la siguiente imagen:



**ILUSTRACIÓN 4. DIAGRAMA FUNCIONAMIENTO SAGEMAKER**

Amazon SageMaker se puede resumir en las fases clave de construcción (build), prueba y ajuste (test & tune), y despliegue (deploy) de modelos de machine learning. A continuación, se detalla cada una de estas fases:

✓ Construir (Build):

**Entrenamiento de Modelos:** En la fase de construcción, los científicos de datos y desarrolladores utilizan SageMaker para entrenar sus modelos de machine learning. Esto implica proporcionar conjuntos de datos, seleccionar algoritmos de machine learning y ajustar los parámetros del modelo. SageMaker facilita el proceso de entrenamiento, permitiendo la escalabilidad horizontal para manejar grandes volúmenes de datos y realizar experimentos eficientes.

✓ Probar y Ajustar (Test & Tune):

**Optimización de Modelos:** Una vez que se ha construido un modelo, es esencial evaluar su rendimiento. SageMaker proporciona herramientas para probar y ajustar el modelo, incluyendo la capacidad de realizar ajustes finos en los hiperparámetros para mejorar la precisión y eficacia del modelo. Esta fase ayuda a garantizar que el modelo cumpla con los requisitos y objetivos establecidos.

✓ Desplegar (Deploy):

**Implementación de Modelos:** La fase final es la implementación del modelo en entornos de producción. SageMaker permite desplegar modelos de machine learning de manera rápida y sencilla, proporcionando una infraestructura escalable y gestionada para ejecutar inferencias en tiempo real. Puedes desplegar modelos en instancias de SageMaker, servidores web, o incluso en dispositivos edge con AWS IoT Greengrass.

Resumir SageMaker en estas tres fases refleja el flujo de trabajo típico en el desarrollo de modelos de machine learning, desde la construcción y entrenamiento inicial hasta la evaluación y ajuste continuo, culminando en la implementación efectiva en entornos de producción. Esta estructura simplificada ayuda a los usuarios a comprender y aplicar de manera eficiente las capacidades completas de SageMaker a lo largo de todo el ciclo de vida de los modelos de machine learning.

- **AWS Deep Learning AMIs:**
  - **Imágenes de Máquinas Virtuales Preconfiguradas:** AWS proporciona Amazon Machine Images (AMIs) preconfiguradas con bibliotecas y entornos necesarios para el desarrollo de deep learning.
- **Amazon Comprehend:**
  - **Procesamiento de Lenguaje Natural (NLP):** Este servicio permite analizar texto para extraer información clave, como sentimientos, entidades y relaciones.
- **Amazon Polly:**
  - **Generación de Voz:** Polly convierte texto en discurso (text-to-speech), permitiendo la creación de aplicaciones con capacidad de hablar.
- **Amazon Rekognition:**
  - **Visión por Computadora:** Ofrece capacidades de reconocimiento de imágenes y vídeos para identificar objetos, caras y texto, así como analizar contenido visual.
- **AWS DeepLens:**
  - **Dispositivo de Aprendizaje Profundo:** Es una cámara con capacidad de aprendizaje profundo diseñada para ayudar a los desarrolladores a aprender y construir modelos de ML en el borde de la red.

Estos son solo algunos ejemplos de la amplia gama de servicios que AWS ofrece en el ámbito de Machine Learning. La plataforma busca hacer que el desarrollo y despliegue de modelos de ML sea accesible, escalable y eficiente para una variedad de aplicaciones y casos de uso.

# MICROSOFT AZURE

## ¿Qué es Microsoft Azure?

Según la página oficial de Microsoft, Azure es una plataforma basada en servidores (nube) que tiene más de 200 productos y servicios para poder solucionar los problemas del día a día. Se puede crear y ejecutar aplicaciones creadas en múltiples nubes con herramientas a escoger de una amplia variedad ofrecidas por Azure.

Una explicación de qué hacen los Azure web Services se puede encontrar tanto en material académico [ResearchGate](#) (2021, IJSRMS) como en diversos sitios aficionados. Siendo el factor común que Azure tiene las mismas capacidades computacionales que Google Cloud. Azure es la segunda plataforma más grande de cloud computer. Dentro de los servicios que ofrece Azure, existe el ML dentro de la nube.

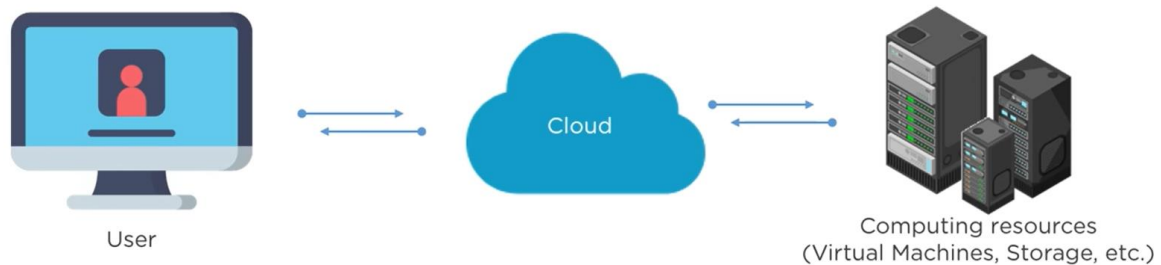
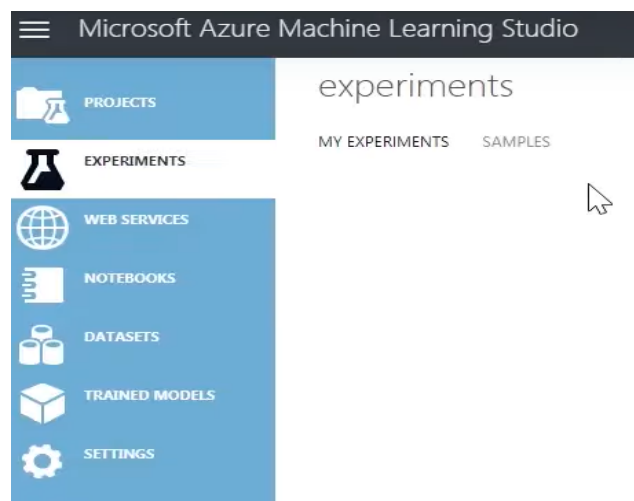


ILUSTRACIÓN 3. DIAGRAMA DE FUNCIONAMIENTO AZURE

## Cloud ML con Azure

Azure cuenta con una plataforma específica para poder realizar Machine Learning con los recursos computacionales de la nube y esta se llama "Microsoft Azure Machine Learning Studio". En este ambiente se pueden crear proyectos de ML que serán procesados por los servidores de Azure.



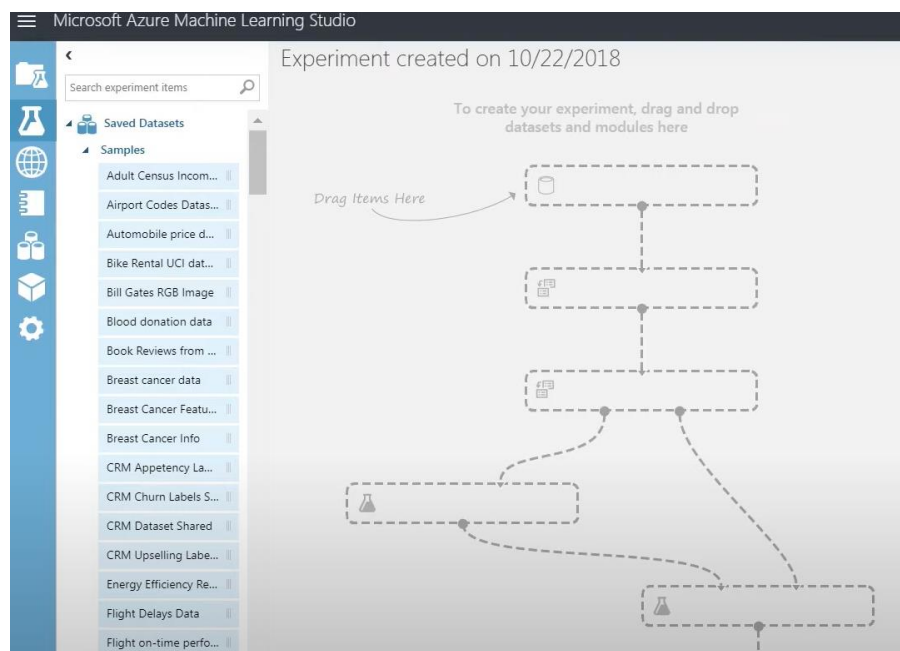


La plataforma cuenta con varias opciones para la creación de programas ML.:

- **Experiments:** Todos los programas usados para crear el modelo
- **Web Services:** Permite subir los experimentos a la nube para ser corrido en cualquier ordenador conectado a la misma
- **Notebooks:** Permite tener código
- **Datasets:** Son todos los datos que se utilizan para el modelo
- **Trained Models:** En este se almacenan todos los modelos de ML entrenados

## Creación de un modelo ML

Los experimentos permiten crear un flujo en el que se interconectan los bloques utilizados para crear el programa. En este ejemplo [particular](#), se busca crear un modelo que permita calcular el salario de una persona que es nueva para la base de datos.



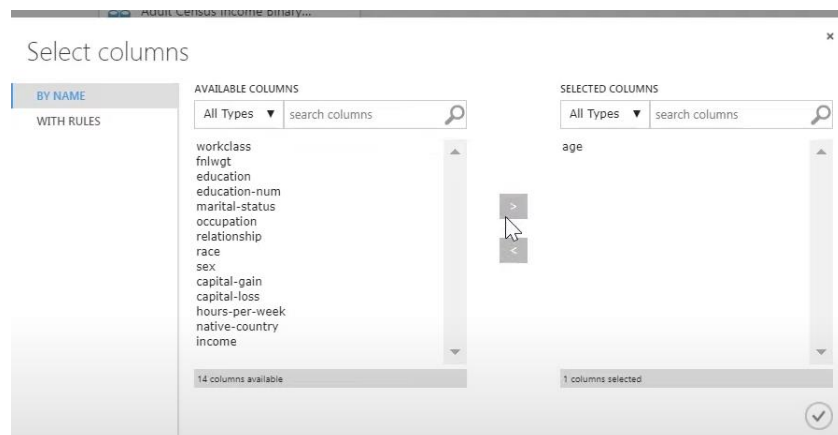
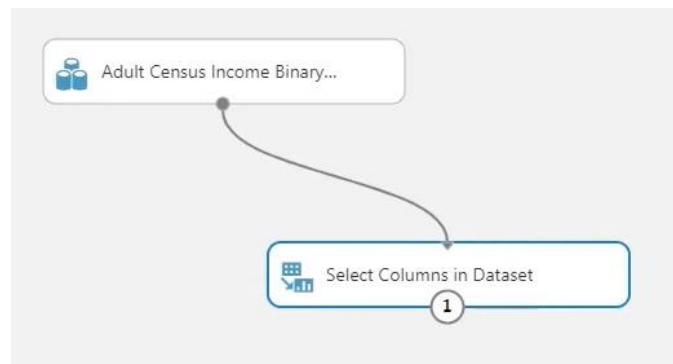
Para ello lo primero que se hace es ingresar los datos con los que se van a trabajar (que previamente deberían estar ingresados en Datasets). Este set de datos puede ser visualizado en forma tabular para ver el tipo de dato, la distribución y la cantidad de datos.

Income Prediction > Adult Census Income Binary Classification dataset > dataset

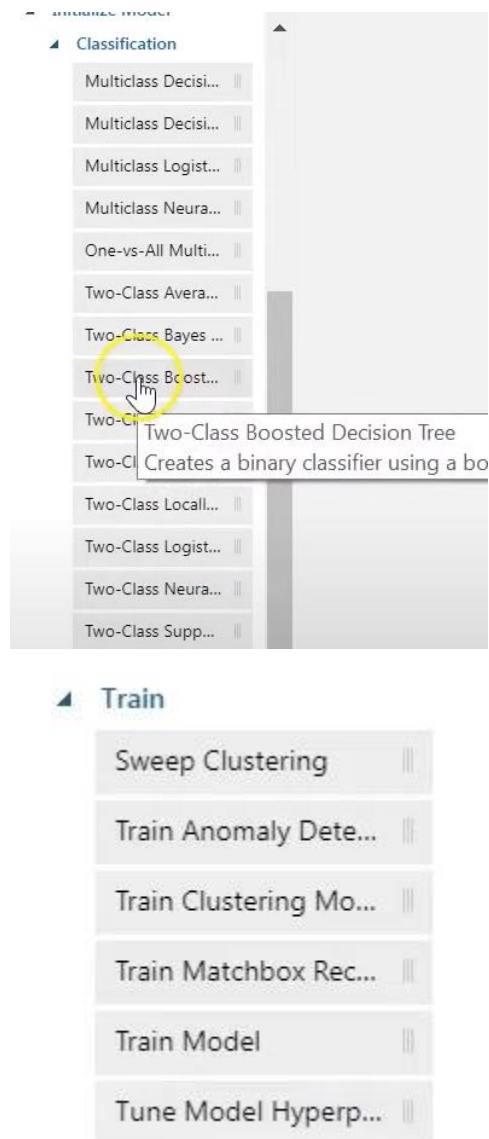
rows: 32561, columns: 15

workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income
State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K
Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K

El preprocesado se puede realizar por medio de tareas o bloques que se agregan al flujo de datos. En este caso existen diversas herramientas que permiten realizar este preprocesado solo tienen que ser agregadas al modelo



Una vez que se tienen los datos preprocesados se pueden aplicar los algoritmos con los que ya cuenta la plataforma. Por ejemplo, en este caso se quiere realizar una clasificación por lo que se ven los algoritmos de clasificación disponibles y se incluye el que se prefiera usar para el data set elegido. De la misma forma el entrenamiento y los parámetros de este, son modelados por medio de un bloque de los diferentes tipos de entrenamiento ofrecidos por Azure.

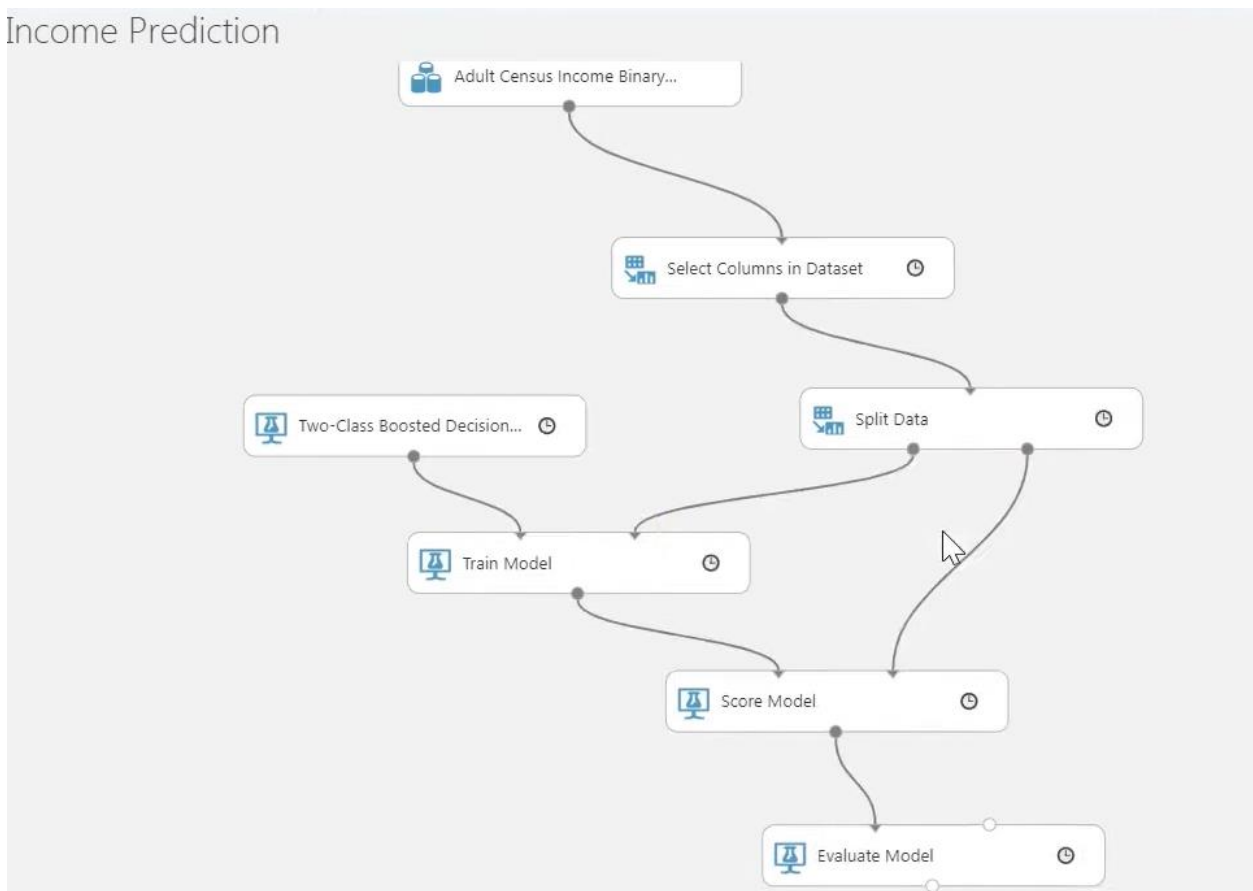


Además de esto Azure permite poner bloques tanto de división del set de datos entre entrenamiento y evaluación; así como también permite definir la métrica que se utilizará para evaluar el rendimiento del algoritmo.

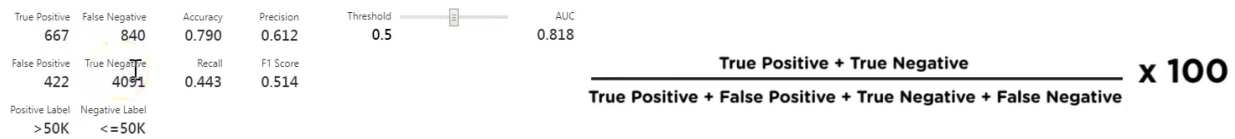


Con esto se concluyen todas las etapas del entrenamiento automático. Y cada una de ellas cuenta con un módulo diferente dentro del ambiente de Microsoft Azure. Desde el ingreso de datos, pasando por el preprocesado y preparación de los datos; hasta aplicar el algoritmo y evaluar su funcionalidad con los datos de entrenamiento y de prueba.

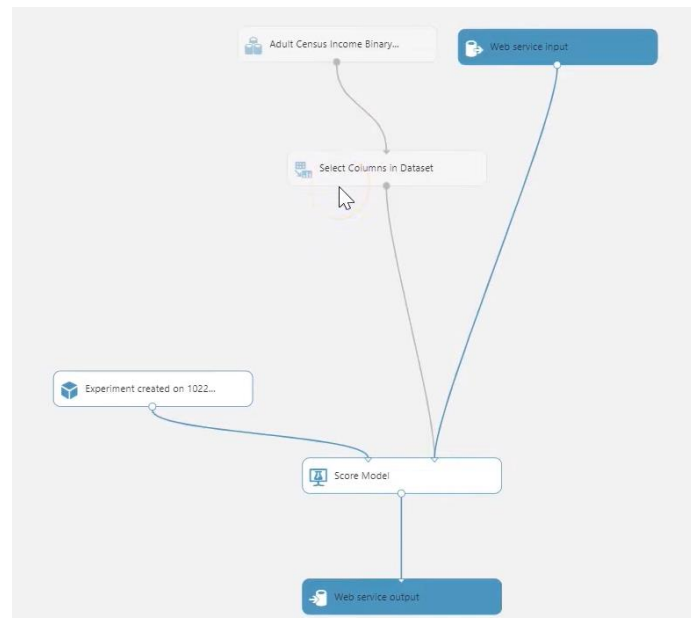
## Income Prediction



Finalmente se procede a correr el modelo y a evaluarlo con el data set ingresado. En este caso se utilizó la métrica de matriz de confusión para el experimento y se obtiene un dato de Accuracy del 79% lo cual es aceptable para el modelo que busca clasificar los salarios de las personas.



Finalmente, otra ventaja que ofrece Azure como un ML cloud Service es que no solo se entrenó utilizando programas y recursos ajenos, sino que puede ser usado desde cualquier dispositivo con acceso a la red. Este seguirá usando el modelo que se creó con el entrenamiento del data set inicial.



De esta forma lo único que queda es ingresar datos externos y utilizará el modelo que le brindamos inicialmente para generar predicciones basándose en las entradas que se le proporcionen.

