



Alfavalue

Módulo 1: Entendiendo el Ciclo de Ciencia de Datos

Agenda

01

Introducción al
proceso de
Ciencia de Datos

02

Opciones de
Ciencia de Datos
en Azure

03

Introducción a
Azure NoteBooks

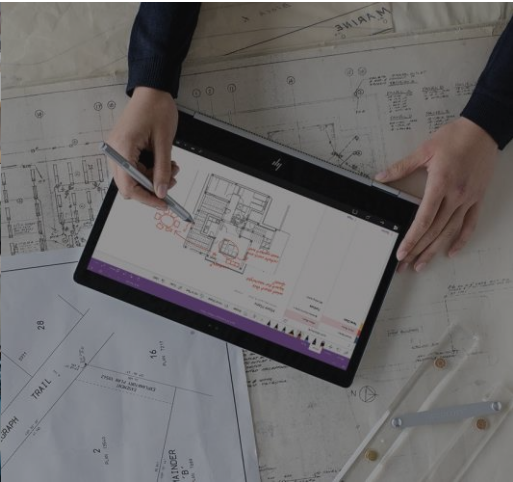
04

Ejercicio: Creando
un Modelo de
Machine Learning
usando Azure
NoteBooks



Lección 1

Introducción al proceso de Ciencia de Datos



Casos de Uso Ciencia de Datos



Aumentar ventas al predecir la demanda de un producto para que el inventario esté listo.



Redu reingresos hospitalarios identificando pacientes de alto riesgo.



Identificar los bienes asegurados con alto riesgo de daños por tormenta.

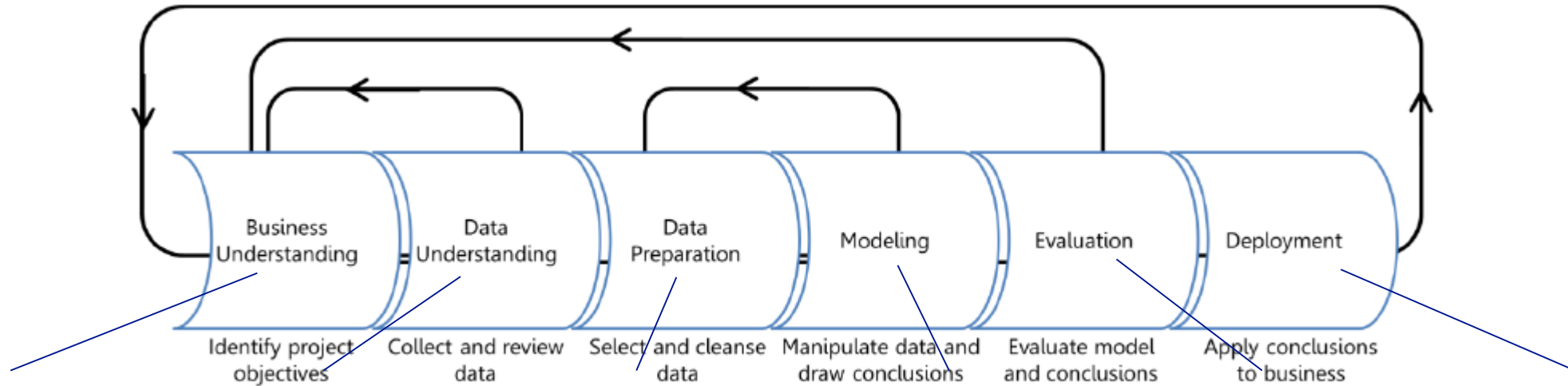


Análisis de imágenes médicas para identificar tumores cancerosos.



Identifique las mejores opciones de tratamiento para pacientes con cáncer.

El Proceso de la ciencia de datos



Problema:
La compañía pierde ventas porque el inventario está agotado.

¿podemos predecir la demanda?

científicos de datos exploran los datos disponibles para encontrar patrones y relaciones con el problema comercial, es decir, ¿qué impulsa la demanda?

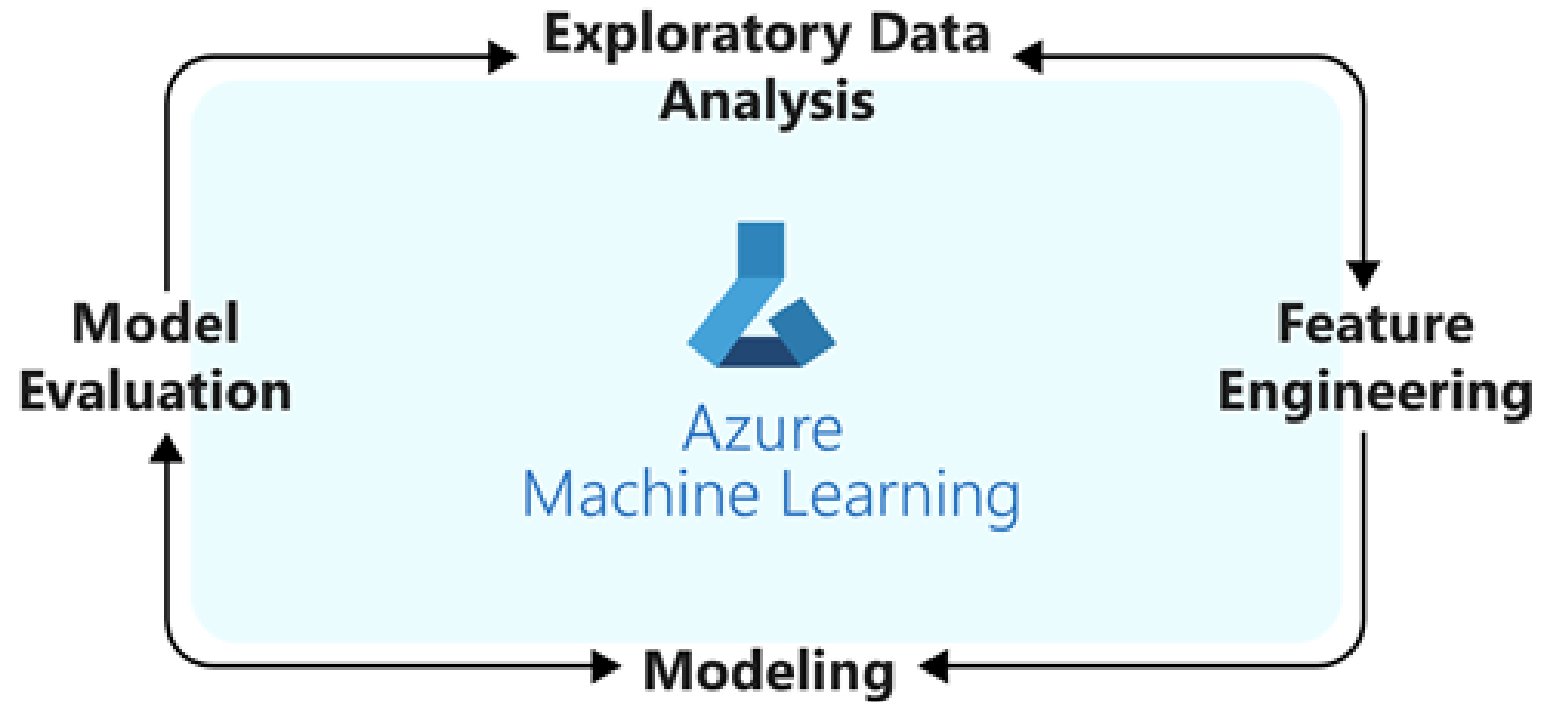
científicos de datos limpian y transforman los datos según sea necesario para soportar el modelado. Algunas columnas pueden necesitar ser agregadas y otras reformateadas.

Las características, los atributos que se pueden usar para predecir el resultado, se identifican y extraen, los modelos potenciales de aprendizaje automático se seleccionan y entrenan.

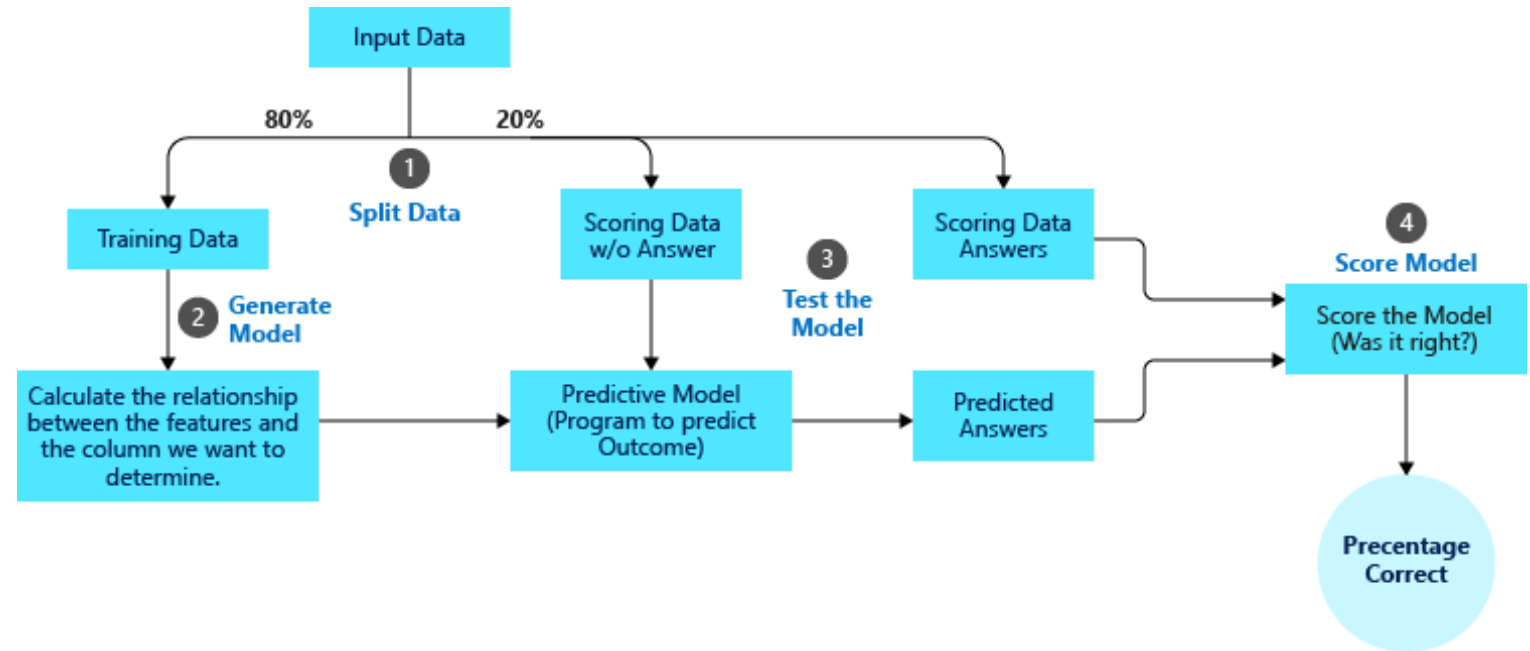
Los modelos entrenados se evalúan y comparan para encontrar el que tenga el mejor rendimiento, es decir, el que predice las ventas de manera más confiable.

El modelo con mejor rendimiento se implementa en un entorno donde las aplicaciones pueden usarlo para hacer predicciones.

El Ciclo de Entrenamiento de Modelos



Entrenamiento de Modelos: Pasos



- Divide los datos en Entrenamiento y Pruebas.
- Generar (entrenar) el modelo.
- Prueba el modelo. Use los datos de prueba para hacer predicciones con el modelo.
- Califique el modelo, es decir, compare los valores pronosticados con los reales.

Entrenamiento de modelo: evaluación de resultados



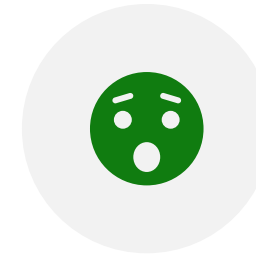
VERDADERO POSITIVO: LA CANTIDAD DE VECES QUE UN MODELO PREDICE VERDADERO (O SÍ) CUANDO EN REALIDAD ES SÍ



VERDADERO NEGATIVO: LA CANTIDAD DE VECES QUE UN MODELO PREDICE FALSO (O NO) CUANDO EN REALIDAD ES FALSO



FALSO NEGATIVO: LA CANTIDAD DE VECES QUE UN MODELO PREDICE FALSO CUANDO EN REALIDAD ES VERDADERO



FALSO POSITIVO: LA CANTIDAD DE VECES QUE UN MODELO PREDICE VERDADERO CUANDO EN REALIDAD ES FALSO

Roles Especializados en Ciencia de datos



Analista de Negocio



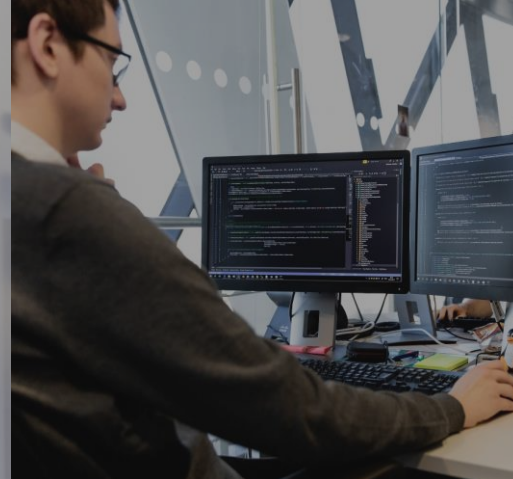
Ingeniero de Datos (Limpieza y estandarización)



Desarrollador

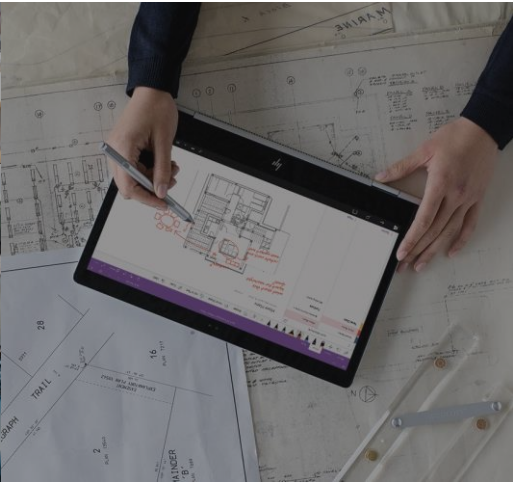


Científico de datos



Lección 2

Seleccionando el servicio de Ciencia de Datos de Azure



Opciones de Ciencia de datos



Azure Machine
Learning Service



Azure Machine
Learning Studio



Azure Databricks



Azure Data Science
Virtual Machine



SQL Server
Machine Learning
Service

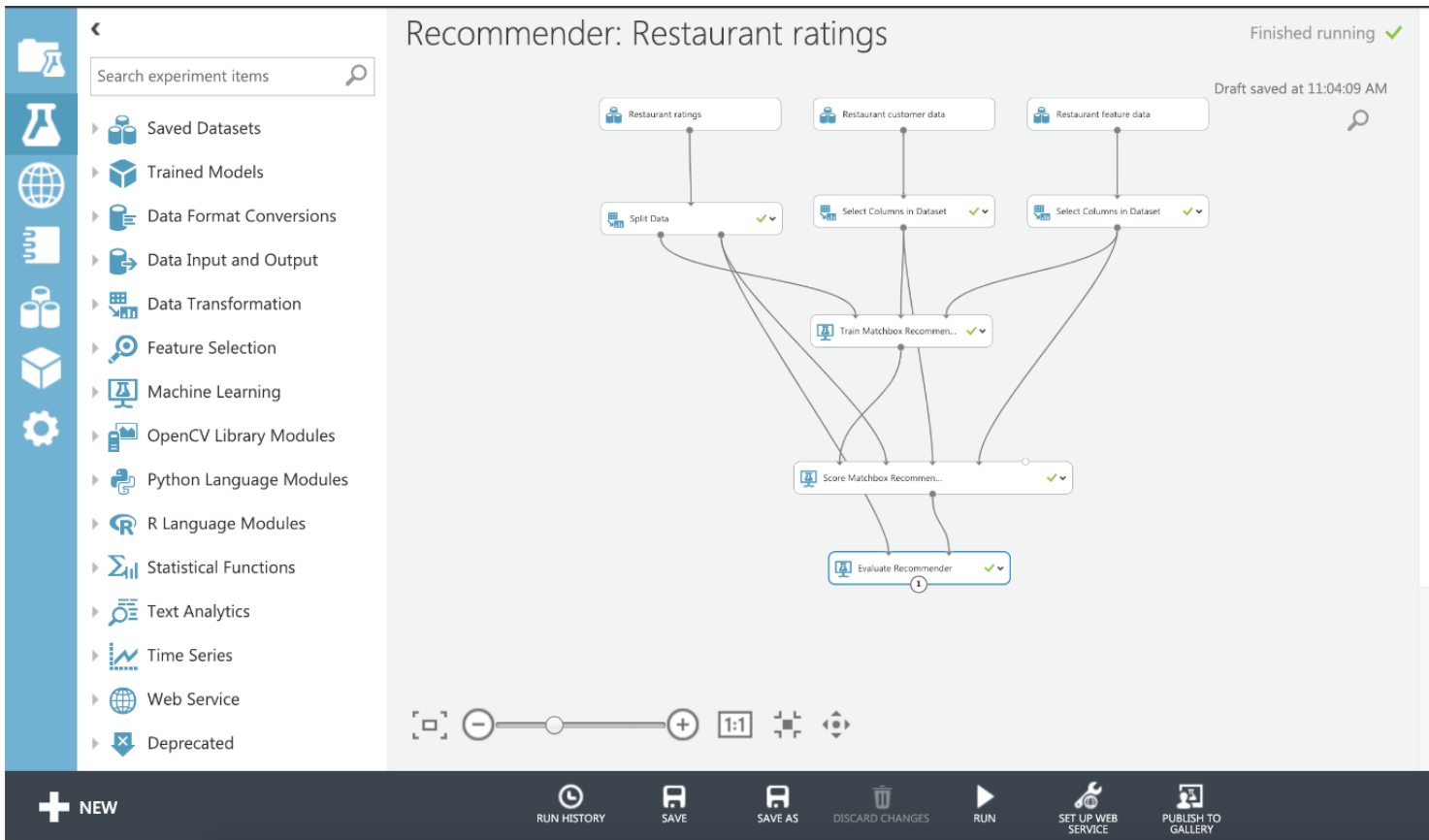
Azure Data Science Virtual Machines

- Máquina virtual personalizada
- Herramientas populares de ciencia de datos preinstaladas y preconfiguradas
- Diseñadas para ayudar a analistas, ingenieros de datos, científicos de datos y desarrolladores.
- Código de muestra para usar los servicios relacionados con la ciencia de datos de Azure



The background image shows a person's hand pointing at a computer monitor. The monitor displays a code editor with HTML and CSS code. Another monitor is visible in the background, and a laptop keyboard is in the foreground. The text is overlaid on the image in a white, bold, sans-serif font.

Práctica 1: Provisionando un Azure Data Science Virtual Machines



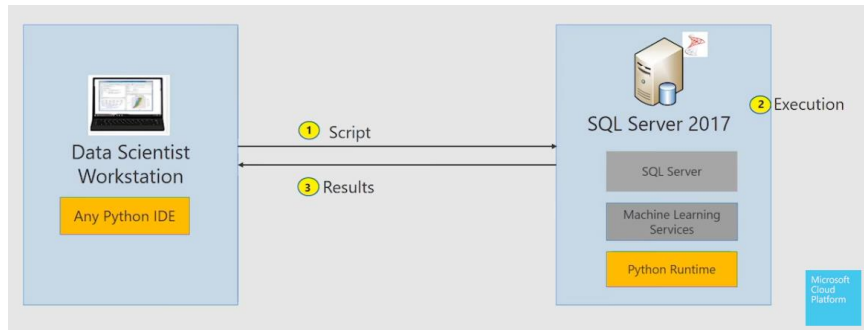
- Herramienta de Arrastrar y soltar para entrenar y desplegar modelos de ML
- Ambiente Completo de Machine Learning
- Ideal para aprender y para Científicos de datos Junior

Azure Machine Learning Studio

¿Por qué SQL Server / Machine Learning Server?

- Python y R de código abierto no se escalan.
- Preocupaciones de seguridad.
- Operacionalización
- Integración y soporte



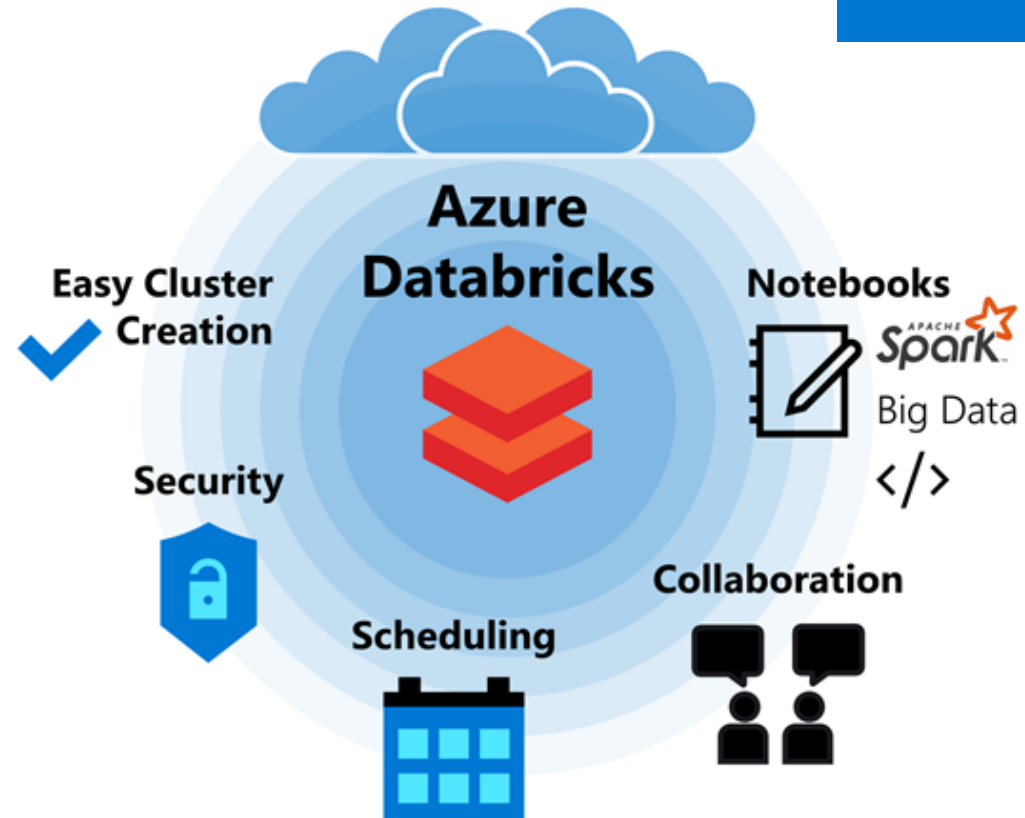


SQL Server/Machine Learning Server

- Escalable en una sola máquina
- Python y código R en procedimientos almacenados.
- Las características de seguridad de SQL Server.

Azure Databricks

- Escala masiva con Spark
- Portal amigable para el usuario
- Escala dinámica
- Seguro
- Herramientas de ciencia de datos



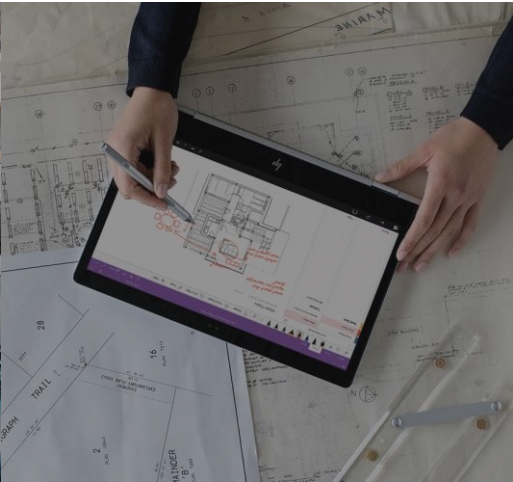
Azure Machine Learning service

- Administración de modelos
- Entrenamiento de modelo
- Selección de modelo
- Ajuste de hiperparámetros
- Selección de características
- Evaluación modelo



Lección 3

Introducción a Ciencia de Datos usando Azure NoteBooks



Qué son Jupyter Notebooks?



Fuente abierta



Originalmente
escrito para Python
y llamado IPython



Ejecución y
renderizado en
tiempo real.



Muchos idiomas
soportados.



Admite API de Spark
(pySpark, SparkR)



Puede compartir a
través de GitHub,
JupyterHub y Azure.

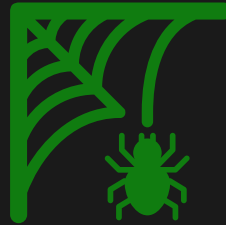


Soporte de visor
incorporado en
GitHub.

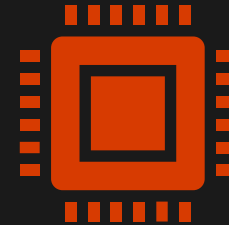
Qué son los Azure Notebooks?



Azure basado en Jupyter Notebooks



Capa gratuita



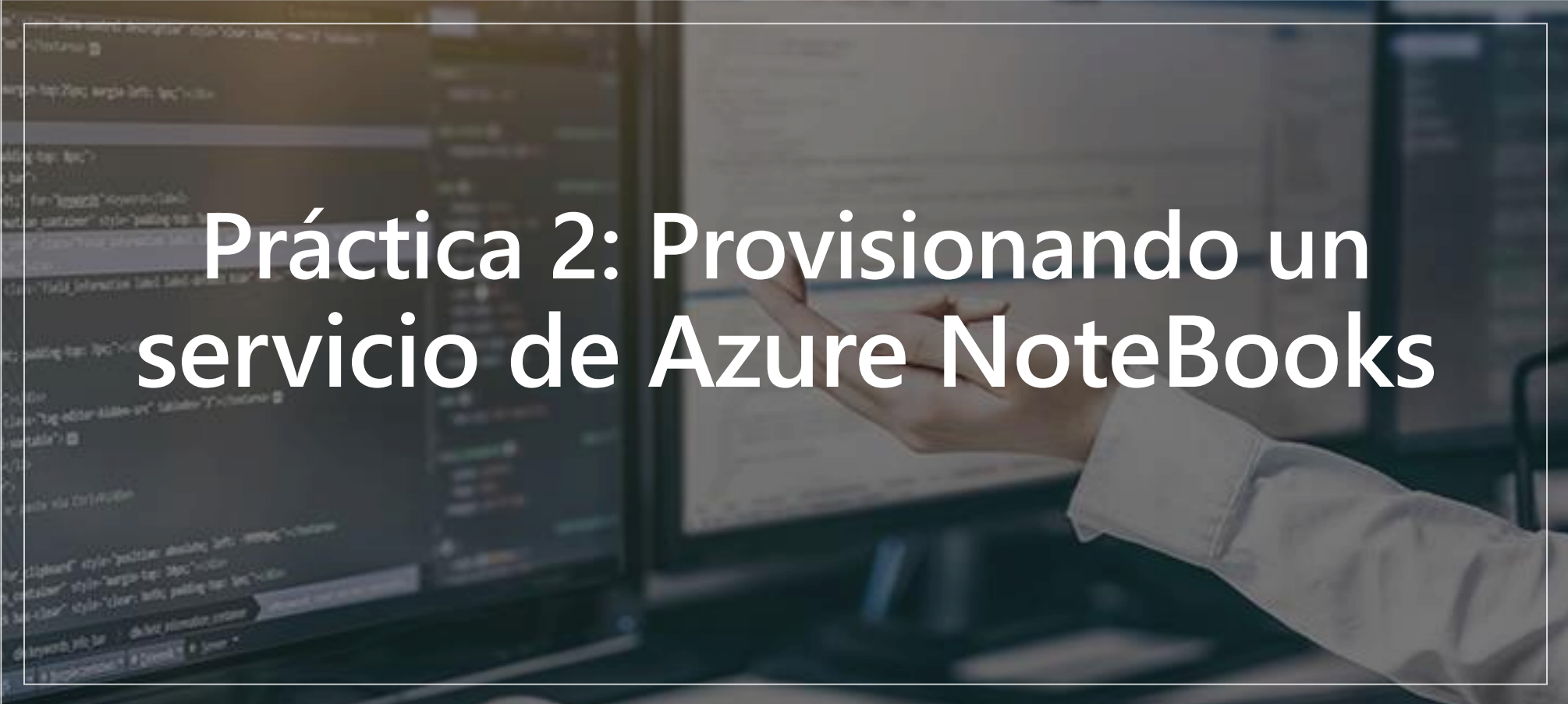
Proyecto listo para usar que enseña cómo usar los datos de Azure y los servicios de inteligencia artificial

Azure Notebook


The screenshot displays the Microsoft Azure Notebook interface. At the top, a dark navigation bar includes 'Microsoft Azure Notebooks', 'Preview', 'My Projects', 'Help', and a user profile 'anon-1oggia'. Below this, a light gray header shows 'Powered by jupyter' and the notebook title 'Introduction to Python (autosaved)'. A blue callout box labeled 'Cell management toolbar' points to the top toolbar, which contains icons for file operations, cell navigation, execution, and output. The main workspace features a menu bar (File, Edit, View, Insert, Cell, Kernel, Azure, Widgets, Help) and a toolbar with icons for saving, adding, deleting, and running cells, along with a 'Markdown' dropdown and a 'Enter/Exit RISE Slideshow' button. A blue callout box labeled 'Rich annotations' points to the text area. The text area contains a section titled 'Comments' with a paragraph explaining Python comment syntax and a code cell. A blue callout box labeled 'Cells contain code or annotations' points to the code cell. The code cell is labeled 'In [1]:' and contains the following code:

```
# this is the first comment
spam = 1 # and this is the second comment
        # ... and now a third!
text = "# This is not a comment because it's inside quotes."
print(text)

# This is not a comment because it's inside quotes.
```



Práctica 2: Provisionando un servicio de Azure NoteBooks



Lab 1: Entrenar un modelo de Clasificación en Azure Notebook
