Learn how to find the best classification model with automated machine learning (AutoML). You'll use the Python SDK (v2) to configure and run an AutoML job.

**Learning objectives**

In this module, you'll learn how to:

* Prepare your data to use AutoML for classification.
* Configure and run an AutoML experiment.
* Evaluate and compare models.

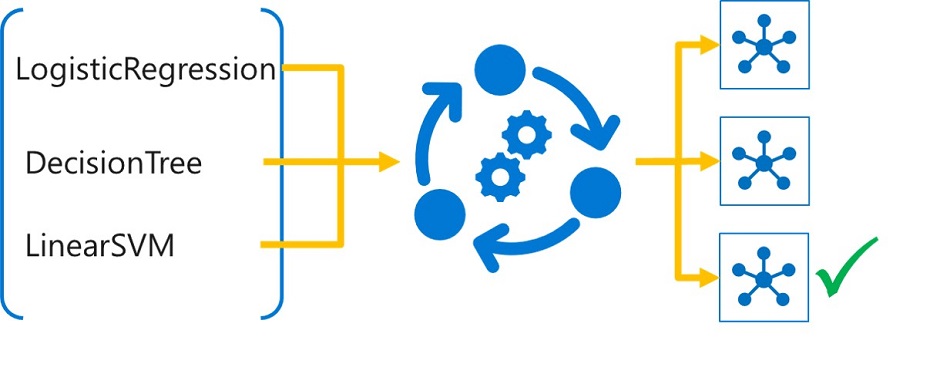
-- **Resumen de AutoML en Azure Machine Learning**

AutoML facilita encontrar el modelo de clasificación con mejor rendimiento al automatizar el proceso de prueba y evaluación de diferentes configuraciones y algoritmos. En Azure Machine Learning, puedes configurar un experimento de AutoML usando la interfaz visual del estudio, la línea de comandos de Azure (CLI) o el SDK de Python.

**Ideas principales**:

* **AutoML**: Automatiza la búsqueda del mejor modelo, ahorrando tiempo en la selección de algoritmos y configuración.
* **Opciones de uso**: Puedes usar AutoML para tareas de clasificación, regresión, pronóstico, clasificación de imágenes y procesamiento de lenguaje natural (NLP).
* **Configuración preferida por data scientists**: El SDK de Python ofrece flexibilidad y control, permitiendo personalizar los experimentos.

Como científico de datos, el SDK de Python suele ser la opción preferida para configurar experimentos de AutoML.



-- **Resumen: Preparación de datos y configuración de featurización en AutoML (Azure Machine Learning)**

Para ejecutar un experimento de AutoML, es crucial preparar y estructurar tus datos adecuadamente.

**Pasos principales**:

1. **Creación de un activo de datos**:
   * Debes crear un "data asset" en Azure Machine Learning usando un archivo MLTable que incluya el esquema de los datos. Esto permite que AutoML interprete correctamente el formato de los datos.

from azure.ai.ml.constants import AssetTypes

from azure.ai.ml import Input

my\_training\_data\_input = Input(type=AssetTypes.MLTABLE, path="azureml:input-data-automl:1")

1. **Transformaciones automáticas de AutoML**:
   * **Escalado y normalización**: AutoML aplica automáticamente técnicas de escalado y normalización para evitar que ciertas características numéricas dominen el proceso de entrenamiento.
   * **Featurización opcional**: AutoML permite aplicar automáticamente técnicas de preprocesamiento, como imputación de valores perdidos, codificación de variables categóricas, eliminación de características de alta cardinalidad y generación de nuevas características (ej., separación de partes de fechas).
   * **Personalización**: Puedes personalizar la featurización, por ejemplo, eligiendo un método específico de imputación para ciertas características.
2. **Análisis posterior al experimento**:
   * Al finalizar el experimento, se puede revisar qué métodos de escalado y normalización se aplicaron, y AutoML te notificará si detecta problemas, como valores perdidos o desequilibrios de clases.

Este proceso asegura que los datos estén bien preparados y optimizados para mejorar el rendimiento del modelo de clasificación en AutoML.

-- **Resumen: Ejecución de un experimento de AutoML en Azure Machine Learning**

**1. Configuración y Ejecución del Experimento**

* Utiliza el SDK de Python para configurar y ejecutar experimentos de AutoML.
* Selecciona algoritmos adecuados según la tarea (por ejemplo, clasificación), como Regresión Logística, Árboles de Decisión, Random Forest, Naive Bayes, SVM, XGBoost, entre otros.

**2. Restricción de Algoritmos**

* Puedes bloquear algoritmos específicos si no se adaptan a tus datos o para cumplir con políticas organizacionales.

**3. Configuración de un Experimento de AutoML**

* Texto

  Descripción generada automáticamenteUsa la función automl.classification para configurar un experimento de clasificación:
* Especifica el **primary\_metric**, que es el criterio de rendimiento principal para seleccionar el modelo óptimo.

**4. Establecimiento de Límites de Entrenamiento**

* Para reducir costos y tiempo, configura límites como timeout\_minutes, trial\_timeout\_minutes, max\_trials, y enable\_early\_termination.
* timeout\_minutes: Number of minutes after which the complete AutoML experiment is terminated.
* trial\_timeout\_minutes: Maximum number of minutes one trial can take.
* max\_trials: Maximum number of trials, or models that will be trained.
* Texto

  Descripción generada automáticamenteenable\_early\_termination: Whether to end the experiment if the score isn't improving in the short term.
* Puedes ejecutar múltiples pruebas en paralelo usando un clúster de cómputo.

**5. Configuración de Propiedades de Entrenamiento**

* Permite a AutoML probar combinaciones de featurización y algoritmos.
* Es posible excluir algoritmos específicos y elegir si usar modelos de ensamblaje para mejorar la precisión.

Este enfoque de AutoML facilita la experimentación automática con varios modelos y configuraciones, optimizando el rendimiento sin intervención manual constante.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente**Envío de un Experimento de AutoML en Azure Machine Learning**

-- **Evaluate and compare models**

Para evaluar y comparar modelos tras completar un experimento de AutoML en Azure Machine Learning, puedes revisar los modelos entrenados para identificar el mejor rendimiento.

1. **Explorar los modelos**: En el Azure Machine Learning Studio, selecciona un experimento de AutoML para ver un resumen del mejor modelo en la página de "Resumen" y acceder a todos los modelos en la pestaña "Modelos".
2. **Explorar pasos de preprocesamiento**: AutoML aplica automáticamente guardias de datos (data guardrails) si se habilita la featurización, que verifican aspectos como:
   * Detección de balance de clases.
   * Imputación de valores faltantes.
   * Detección de alta cardinalidad.

Cada guardia de datos muestra un estado:

* + **Passed** (Aprobado): No se detectaron problemas.
  + **Done** (Hecho): Se realizaron cambios que debes revisar.
  + **Alerted** (Alertado): Se detectaron problemas no solucionados y es necesario revisar los datos.

También se aplican técnicas de escalado y normalización automáticamente.

1. **Recuperar la mejor ejecución**: Los modelos están ordenados por el rendimiento en la métrica primaria especificada. Puedes personalizar las columnas para incluir otras métricas y facilitar la comparación.
2. **Generar explicaciones de modelos**: Puedes activar la generación de explicaciones en el modelo de mejor rendimiento o en otros modelos para evaluar la importancia de características, ayudando a comprender la interpretabilidad del modelo.

Estas explicaciones son aproximaciones que muestran la influencia de las características en la predicción del modelo, ayudando a obtener una visión más detallada del rendimiento y comportamiento de cada modelo.

