**Membuat model klasifikasi dengan perancang Azure Machine Learning**

Klasifikasi adalah teknik pembelajaran mesin yang diawasi yang digunakan untuk memprediksi kategori atau kelas. Pelajari cara membuat model klasifikasi menggunakan perancang Azure Machine Learning.

1. **Tujuan pembelajaran**

* Melatih dan menerbitkan model klasifikasi dengan perancang Azure Machine Learning

1. **Pengantar**

Klasifikasi adalah contoh teknik pembelajaran mesin yang diawasi tempat Anda melatih model menggunakan data yang menyertakan fitur dan nilai yang diketahui untuk label, sehingga model belajar untuk menyesuaikan kombinasi fitur dengan label. Selanjutnya, setelah pelatihan selesai, Anda dapat menggunakan model terlatih untuk memprediksi label untuk item baru yang labelnya tidak diketahui.

Anda dapat menggunakan perancang Microsoft Azure Machine Learning untuk membuat model klasifikasi menggunakan antarmuka visual seret dan lepas, tanpa perlu menulis kode apa pun.

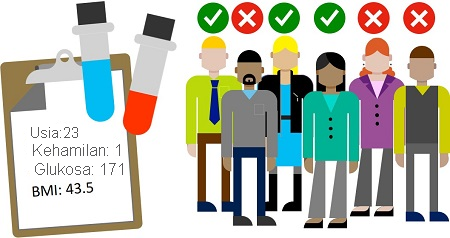
Dalam modul ini, Anda akan mempelajari cara:

* Identifikasi skenario pembelajaran mesin klasifikasi.
* Gunakan perancang Azure Machine Learning untuk melatih model klasifikasi.
* Gunakan model klasifikasi untuk menyimpulkan.
* Menyebarkan dan menguji model klasifikasi.

Untuk menyelesaikan modul ini, Anda memerlukan langganan Microsoft Azure. Jika belum memilikinya, Anda dapat mendaftar untuk uji coba gratis di <https://azure.microsoft.com.>

1. **Mengidentifikasi skenario pembelajaran mesin klasifikasi**

Klasifikasi adalah bentuk pembelajaran mesin yang digunakan untuk memprediksi kategori, atau kelas, item. Misalnya, klinik kesehatan mungkin menggunakan karakteristik pasien (seperti usia, berat badan, tekanan darah, dan sebagainya) untuk memprediksi apakah pasien berisiko terkena diabetes. Dalam hal ini, karakteristik pasien adalah fitur, dan label adalah klasifikasi 0 atau 1, yang mewakili non-diabetes atau diabetes.



Seperti regresi, klasifikasi adalah contoh teknik pembelajaran mesin terawasi di mana Anda melatih model menggunakan data yang menyertakan fitur dan nilai yang diketahui untuk label, sehingga model belajar menyesuaikan kombinasi fitur ke label. Selanjutnya, setelah pelatihan selesai, Anda dapat menggunakan model terlatih untuk memprediksi label untuk item baru yang labelnya tidak diketahui.

1. **Skenario untuk model pembelajaran mesin klasifikasi**

Model pembelajaran mesin klasifikasi digunakan di banyak industri. Beberapa skenarionya adalah:

* Menggunakan data klinis untuk memprediksi apakah pasien akan sakit atau tidak.
* Menggunakan data historis untuk memprediksi apakah bank harus memberikan pinjaman atau tidak.
* Menggunakan karakteristik usaha kecil untuk memprediksi apakah usaha baru akan berhasil.

1. **Apa itu Azure Machine Learning?**

Melatih dan menyebarkan model pembelajaran mesin yang efektif melibatkan banyak pekerjaan, sebagian besar memakan waktu dan padat sumber daya. Azure Machine Learning adalah layanan berbasis cloud yang membantu menyederhanakan beberapa tugas yang diperlukan untuk menyiapkan data, melatih model, dan menyebarkan layanan prediktif. Model pembelajaran mesin klasifikasi dapat dibangun menggunakan Azure Machine Learning.

Yang paling penting, Azure Machine Learning membantu ilmuwan data meningkatkan efisiensi mereka dengan mengotomatiskan banyak tugas yang memakan waktu yang terkait dengan model pelatihan. Hal ini memungkinkan mereka untuk menggunakan sumber daya komputasi berbasis cloud yang menskalakan dengan efektif untuk menghandel data berjumlah besar sambil mengeluarkan biaya hanya jika benar-benar digunakan.

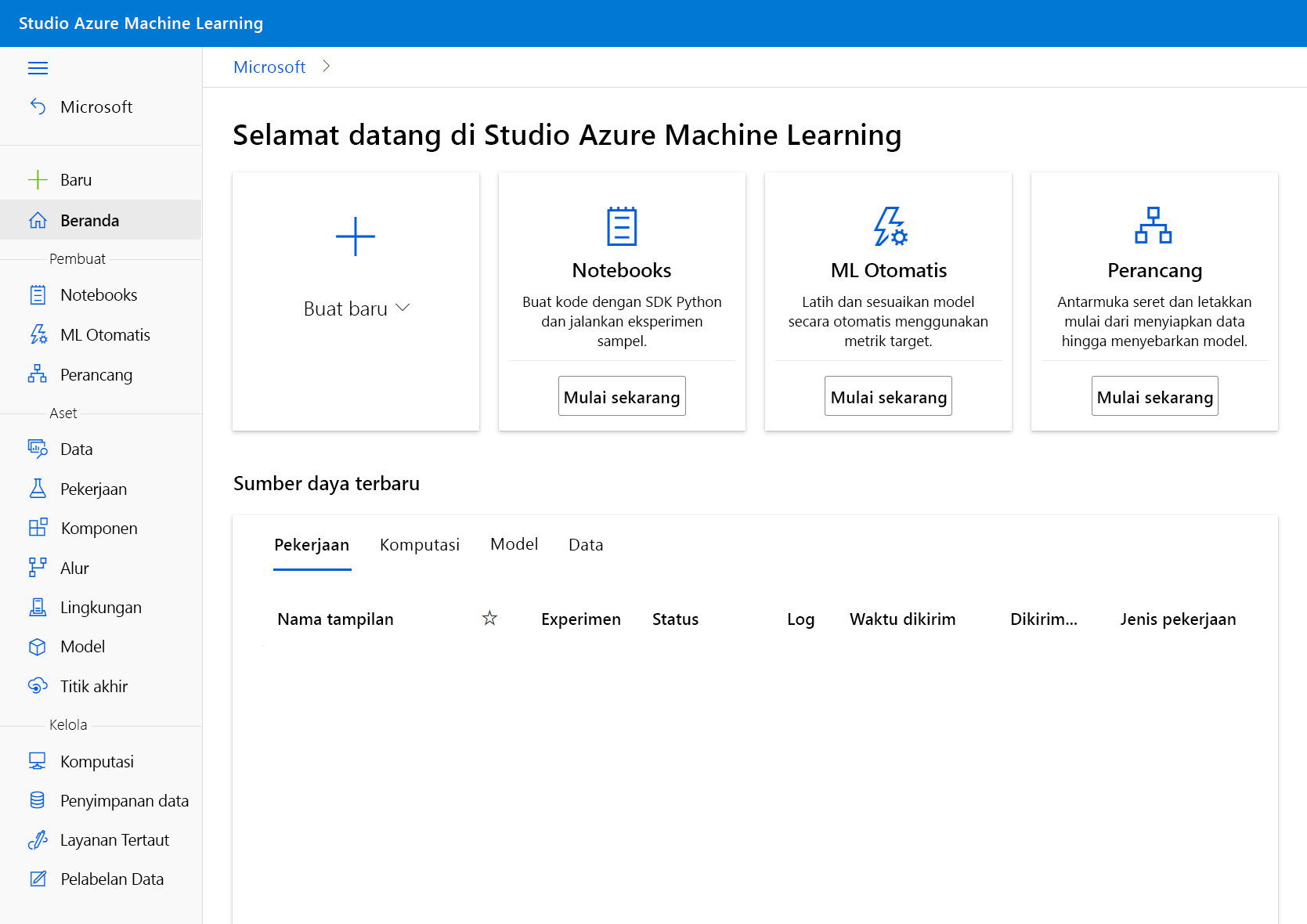
1. **Ruang kerja Azure Machine Learning**

Untuk menggunakan Azure Machine Learning, pertama-tama Anda membuat ruang kerja di langganan Azure Anda. Kemudian, Anda dapat menggunakan ruang kerja ini untuk mengelola data, sumber daya komputasi, kode, model, dan artefak lainnya yang terkait dengan beban kerja pembelajaran mesin Anda.

Setelah Anda membuat ruang kerja Azure Machine Learning, Anda dapat mengembangkan solusi dengan layanan pembelajaran mesin Azure baik dengan alat pengembang atau portal web studio Azure Machine Learning.

1. **Studio Azure Machine Learning**

Studio Azure Machine Learning adalah portal web untuk solusi pembelajaran mesin di Azure. Ini termasuk berbagai fitur dan kemampuan yang membantu ilmuwan data menyiapkan data, melatih model, menerbitkan layanan prediktif, dan memantau penggunaannya. Untuk mulai menggunakan portal web, Anda perlu menetapkan ruang kerja yang Anda buat di portal Azure ke studio Azure Machine Learning.



1. **Komputasi Azure Machine Learning**

Pada intinya, Azure Machine Learning adalah layanan untuk melatih dan mengelola model pembelajaran mesin, di mana Anda memerlukan sumber daya komputasi untuk menjalankan proses pelatihan. Target komputasi adalah sumber daya berbasis cloud tempat Anda dapat menjalankan pelatihan model dan proses eksplorasi data.

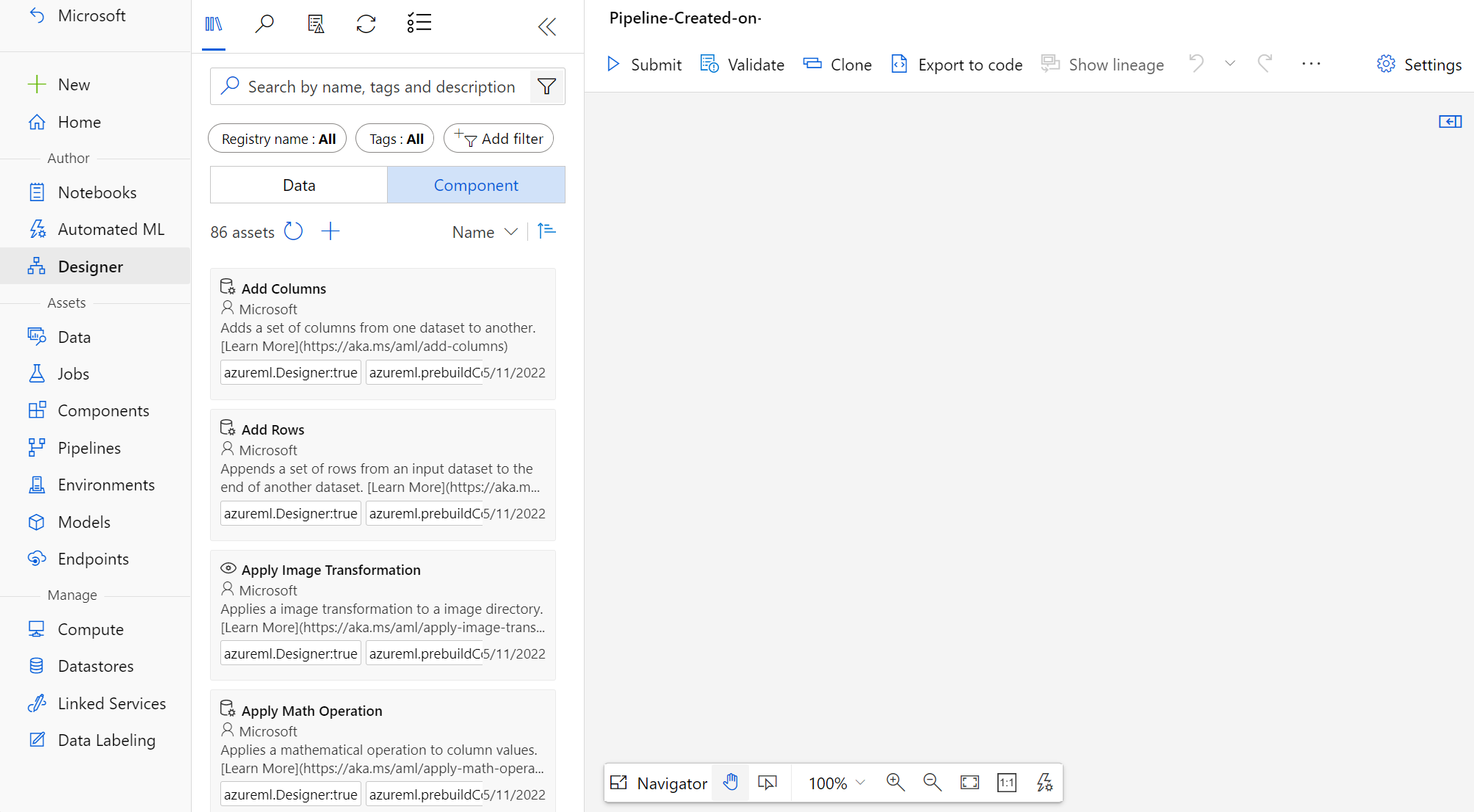
Di studio Azure Machine Learning, Anda dapat mengelola target komputasi untuk aktivitas ilmu data Anda. Ada empat jenis sumber daya komputasi yang dapat dibuat:

* **Instans Komputasi:** Stasiun kerja pengembangan yang dapat digunakan oleh ilmuwan data untuk bekerja dengan data dan model.
* **Kluster Komputasi:** Kluster komputer virtual yang dapat diskalakan untuk pemrosesan kode eksperimen sesuai permintaan.
* **Kluster Inferensi:** Target penyebaran untuk layanan prediktif yang menggunakan model terlatih Anda.
* **Komputasi Terlampir:** Tautan ke sumber daya komputasi Azure yang ada, seperti kluster Virtual Machines atau Azure Databricks.

1. **Apa itu perancang Azure Machine Learning?**

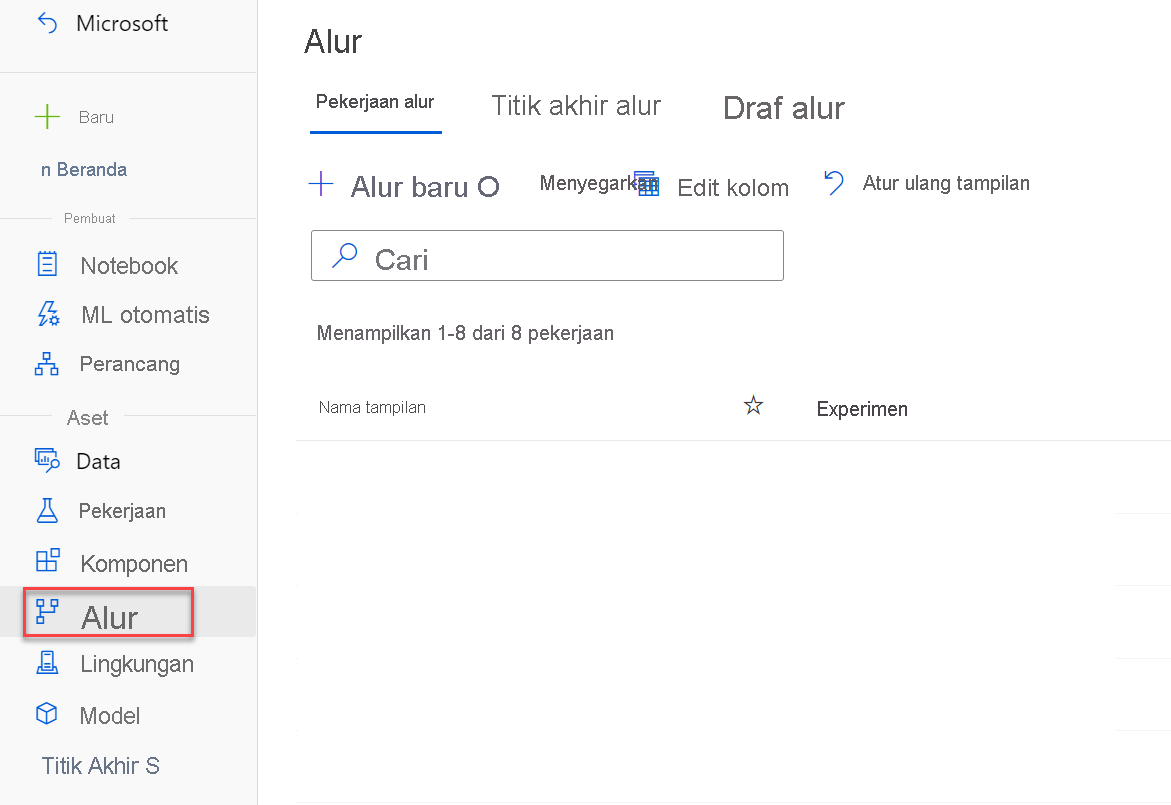
Di studio Azure Machine Learning, terdapat beberapa cara untuk membuat klasifikasi model pembelajaran mesin. Salah satu caranya adalah dengan menggunakan antarmuka visual yang disebut perancang yang dapat Anda gunakan untuk melatih, menguji, dan menyebarkan model pembelajaran mesin. Antarmuka seret dan letakkan menggunakan input dan output yang ditentukan dengan jelas yang dapat dibagikan, digunakan kembali, dan dikontrol versi.

Setiap proyek perancang, yang dikenal sebagai alur, memiliki panel kiri untuk navigasi dan kanvas di sisi kanan Anda. Untuk menggunakan perancang, identifikasi blok penyusun, atau komponen, yang diperlukan untuk model Anda, tempatkan dan sambungkan di kanvas Anda, dan jalankan pekerjaan pembelajaran mesin.



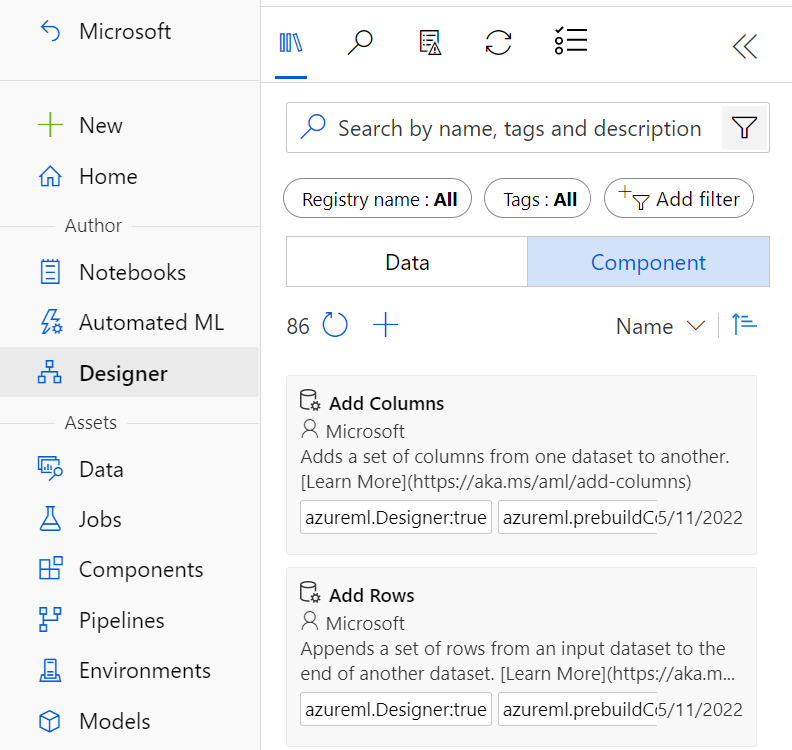
1. **Pipelines**

Alur memungkinkan Anda mengatur, mengelola, dan menggunakan kembali alur kerja pembelajaran mesin yang kompleks di seluruh proyek dan pengguna. Alur dimulai dengan himpunan data tempat Anda ingin melatih model. Setiap kali Anda menjalankan alur, konfigurasi alur dan hasilnya disimpan di ruang kerja Anda sebagai pekerjaan alur.



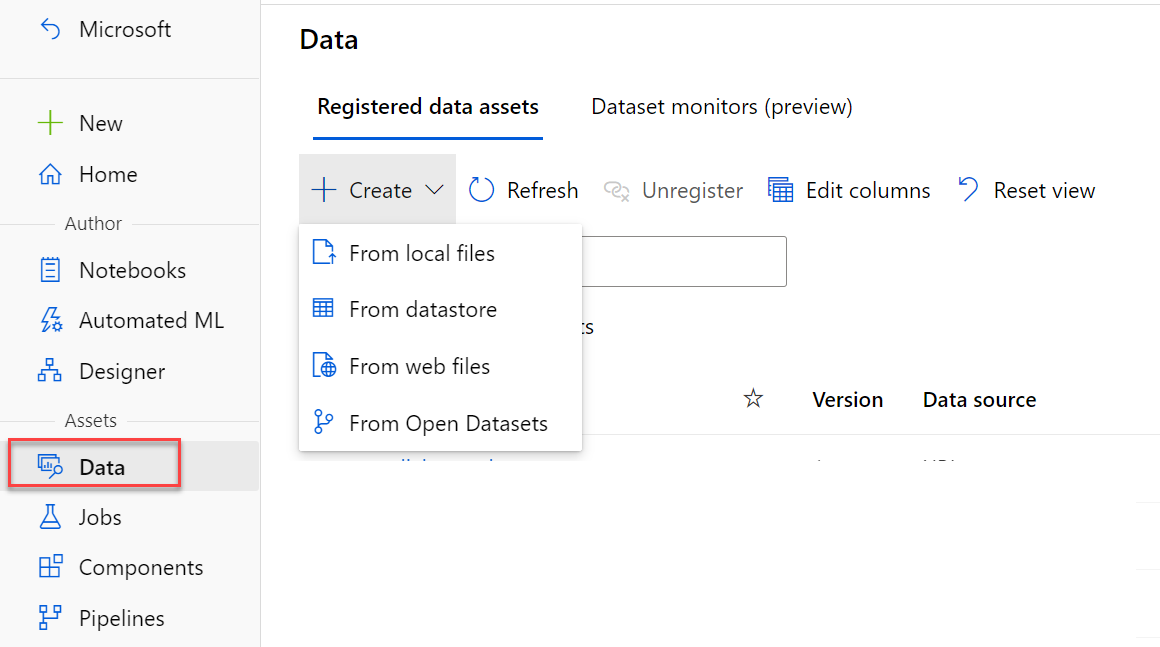
1. **Komponen**

Komponen Azure Machine Learning merangkum satu langkah dalam alur pembelajaran mesin. Anda dapat menganggap komponen sebagai fungsi pemrograman dan sebagai blok penyusun untuk alur Azure Machine Learning. Dalam proyek alur, Anda dapat mengakses aset data dan komponen dari tab **Pustaka Aset** panel kiri.



1. **Himpunan Data**

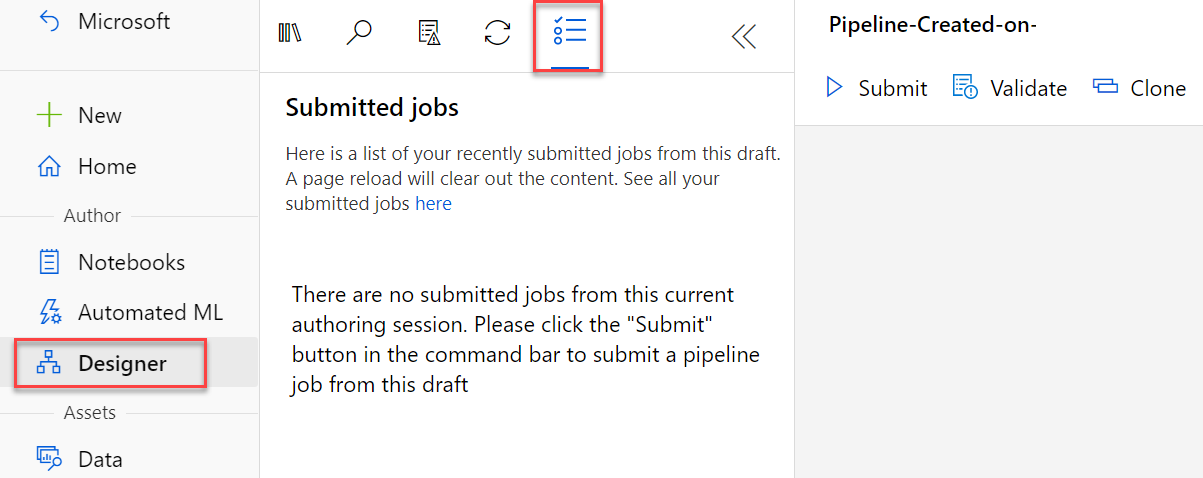
Anda dapat membuat aset data pada halaman **Data** dari file lokal, penyimpanan data, file web, dan Open Datasets. Aset data ini akan muncul bersama dengan himpunan data sampel standar di **Pustaka Aset** perancang.



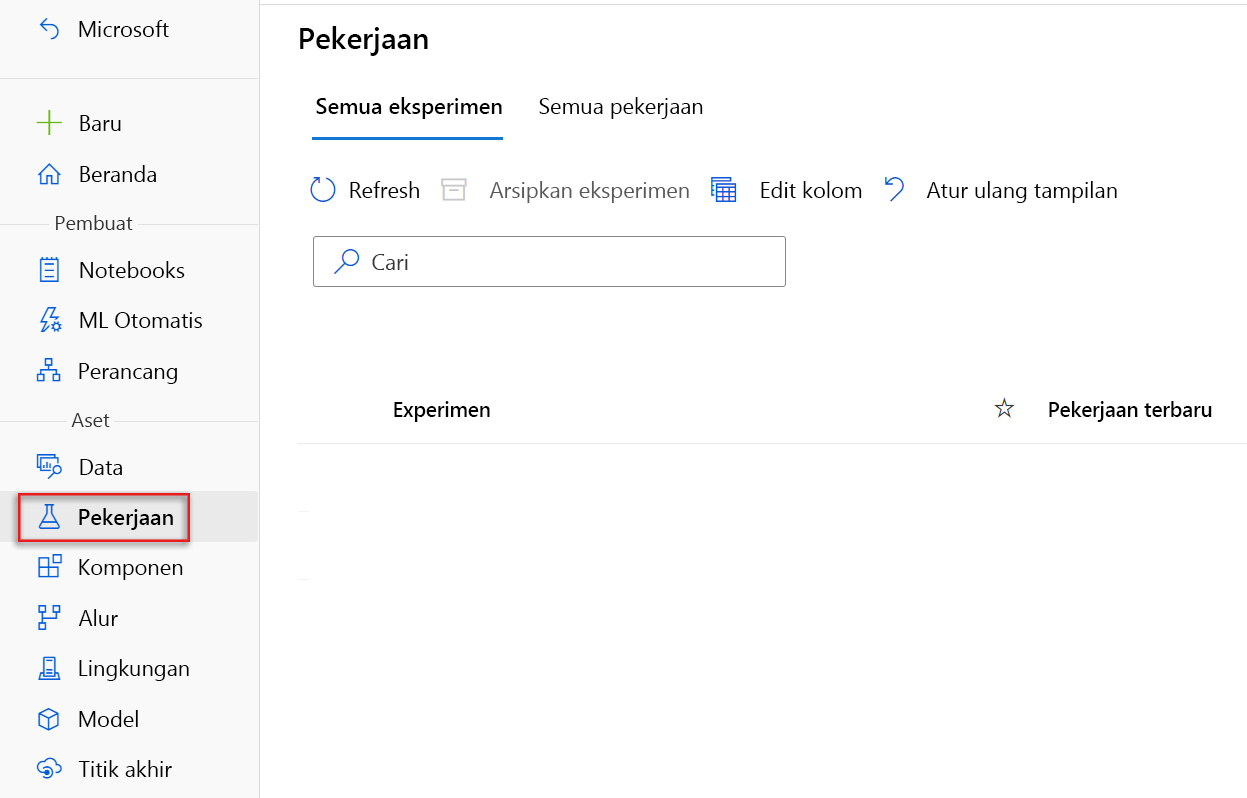
1. **Pekerjaan Azure Machine Learning**

Pekerjaan Azure Machine Learning (ML) menjalankan tugas terhadap target komputasi yang ditentukan. Pekerjaan memungkinkan pelacakan sistematis untuk eksperimen dan alur kerja pembelajaran mesin. Setelah pekerjaan dibuat, Azure ML mempertahankan rekaman eksekusi untuk pekerjaan tersebut. Semua rekaman eksekusi pekerjaan Anda dapat dilihat di studio Azure ML.

Dalam proyek perancang, Anda dapat mengakses status pekerjaan alur menggunakan tab **Pekerjaan yang dikirimkan** di panel kiri.



Anda dapat menemukan semua pekerjaan yang telah dijalankan di ruang kerja pada halaman **Pekerjaan**.



1. **Memahami langkah-langkah untuk klasifikasi**

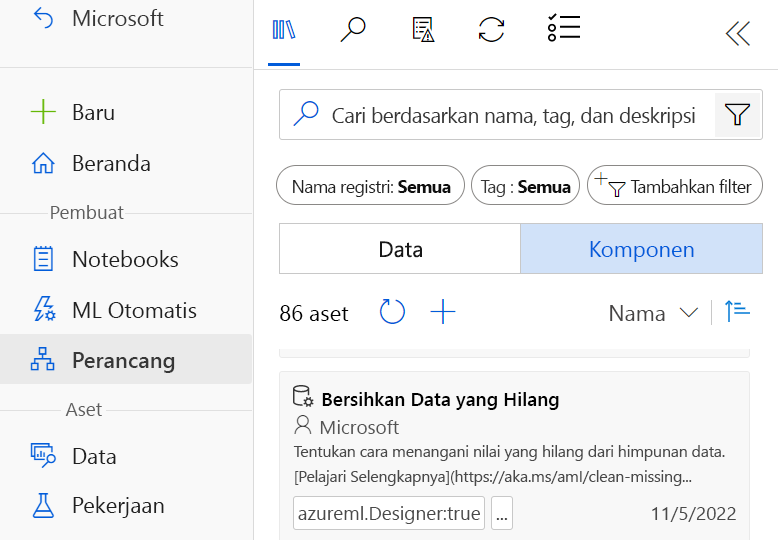
Anda dapat memikirkan langkah-langkah untuk melatih dan mengevaluasi model pembelajaran mesin klasifikasi sebagai:

1. **Menyiapkan data:** Mengidentifikasi fitur dan label dalam himpunan data. Pra-proses, atau membersihkan dan mentransformasi, data seperlunya.
2. **Melatih model**: Membagi data menjadi dua grup, pelatihan, dan set validasi. Latih model pembelajaran mesin menggunakan himpunan data pelatihan. Uji model pembelajaran mesin untuk performa menggunakan himpunan data validasi.
3. **Mengevaluasi performa:** Bandingkan seberapa dekat prediksi model dengan label yang diketahui.
4. **Menyebarkan layanan prediktif:** Setelah melatih model pembelajaran mesin, Anda perlu mengonversi alur pelatihan menjadi alur inferensi real time. Kemudian Anda dapat menyebarkan model sebagai aplikasi di server atau perangkat sehingga orang lain dapat menggunakannya

Mari kita ikuti empat langkah ini saat muncul di Perancang Azure.

1. **Siapkan data**

Perancang Azure Machine Learning memiliki beberapa komponen bawaan yang dapat digunakan untuk menyiapkan data untuk pelatihan. **Komponen-komponen ini memungkinkan Anda membersihkan data, menormalkan fitur, menggabungkan tabel, dan banyak lagi.**



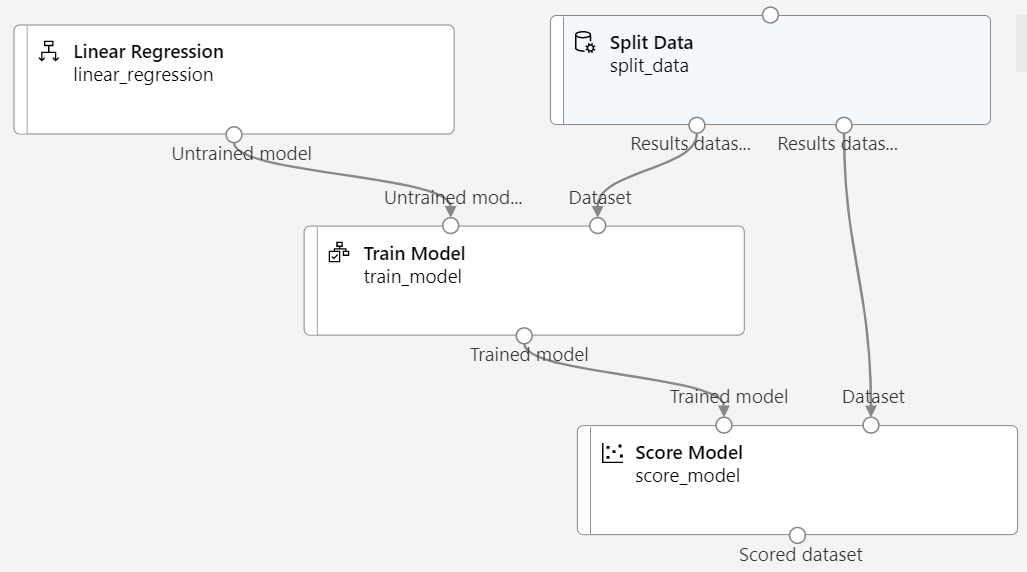
.

1. **Model latihan**

Untuk melatih model klasifikasi, Anda memerlukan kumpulan data yang menyertakan fitur historis, karakteristik entitas yang ingin Anda prediksi, dan nilai label yang diketahui. Label adalah indikator kelas yang ingin Anda latih model untuk memprediksi.

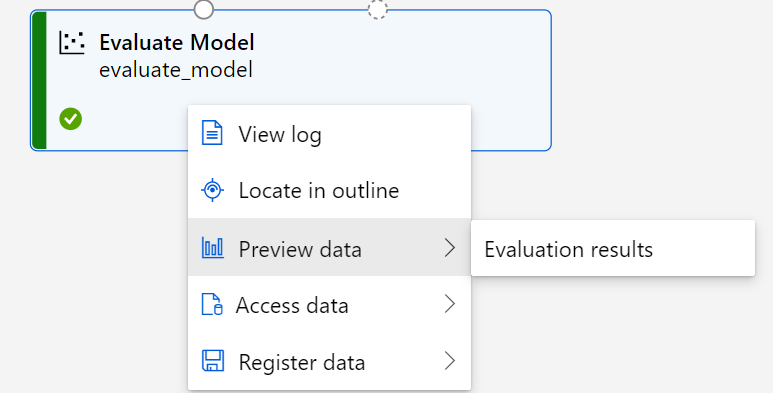
Melatih model menggunakan subset data, sambil menahan beberapa data untuk menguji model terlatih merupakan praktik yang umum. Ini memungkinkan Anda membandingkan label yang diprediksi model dengan label yang diketahui secara aktual dalam himpunan data asli.

Anda akan menggunakan komponen Model Skor **perancang** untuk menghasilkan nilai label kelas yang diprediksi. Setelah Anda menghubungkan semua komponen, Anda akan ingin menjalankan eksperimen, yang akan menggunakan aset data di kanvas untuk melatih dan menilai model.



1. **Evaluasi performa**

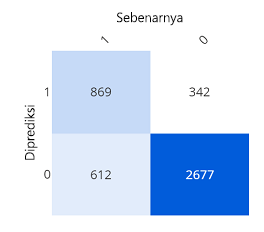
Setelah melatih model, penting untuk mengevaluasi performanya. Ada banyak metrik dan metodologi performa untuk mengevaluasi seberapa baik model membuat prediksi. Anda dapat meninjau metrik evaluasi pada halaman pekerjaan yang telah selesai dengan mengklik kanan komponen **Evaluasi model.**



1. **Matriks kebingungan**

Confusion matrix memperlihatkan kasus ketika nilai prediksi dan aktualnya adalah 1 (dikenal sebagai positif benar) di kiri atas, dan kasus dengan nilai prediksi dan aktualnya adalah 0(negatif benar) di kanan bawah. Sel-sel lain memperlihatkan kasus dengan nilai prediksi dan aktual yang berbeda (positif palsu dan negatif palsu).

Untuk model klasifikasi biner di mana Anda memprediksi salah satu dari dua kemungkinan nilai, matriks konfusi adalah kisi 2x2 yang menunjukkan jumlah nilai yang diprediksi dan aktual untuk kelas 0 dan 1, mirip dengan ini:



Untuk model klasifikasi multi-kelas (yang memiliki lebih dari dua kelas yang mungkin), pendekatan yang sama digunakan untuk menabulasi setiap kemungkinan kombinasi hitungan nilai aktual dan prediksi - agar model dengan tiga kelas yang mungkin akan menghasilkan matriks 3x3 dengan garis sel diagonal yang label prediksi dan aktualnya cocok.

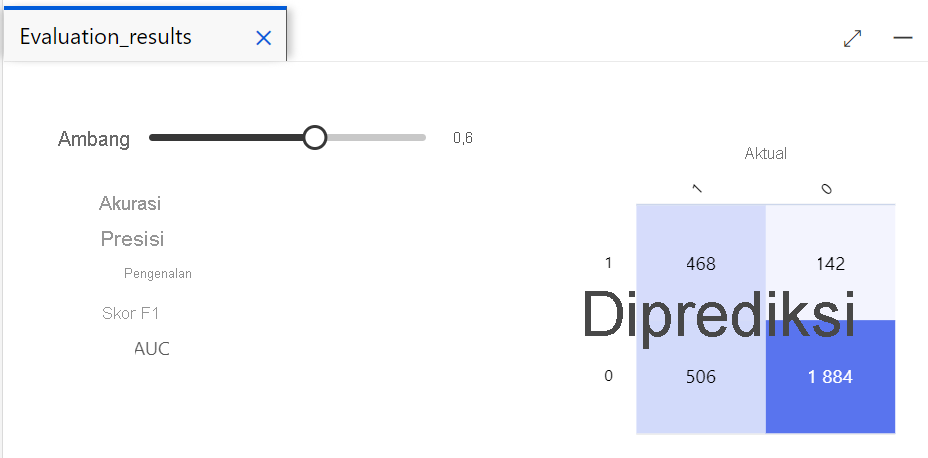
Metrik dapat berasal dari matriks kebingungan meliputi:

* **Akurasi**: Rasio prediksi yang benar (positif benar + negatif benar) terhadap jumlah total prediksi.
* **Presisi**: Pecahan kasus positif yang diidentifikasi dengan benar (jumlah positif benar dibagi dengan jumlah positif benar ditambah positif palsu).
* **Pengenalan**: Pecahan kasus yang diklasifikasikan sebagai positif yang benar-benar positif (jumlah positif benar dibagi dengan jumlah positif benar ditambah negatif palsu).
* **Skor F1**: Metrik keseluruhan yang pada dasarnya menggabungkan presisi dan pengenalan.

1. **Memilih ambang batas**

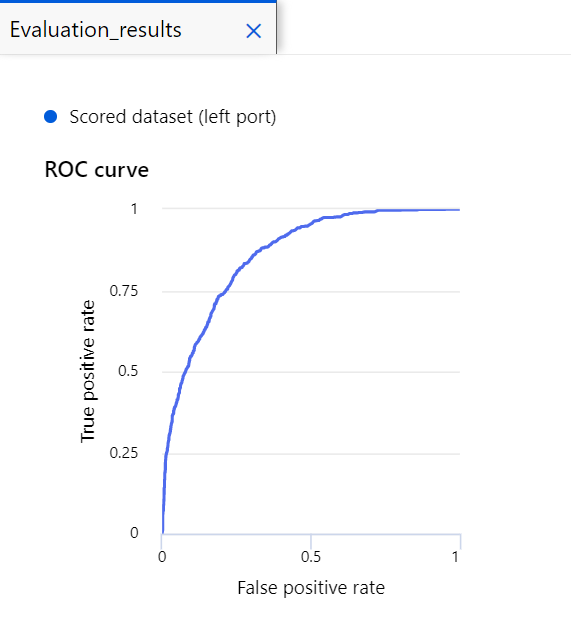
Sebuah model klasifikasi memprediksi probabilitas untuk setiap kelas yang mungkin. Dengan kata lain, model menghitung kemungkinan untuk setiap label yang diprediksi. Dalam kasus model klasifikasi biner, probabilitas yang diprediksi adalah nilai antara 0 dan 1. Secara default, probabilitas yang diprediksi termasuk atau di atas 0,5 menghasilkan prediksi kelas 1, sedangkan prediksi di bawah ambang ini berarti bahwa ada kemungkinan yang lebih besar dari prediksi negatif (ingat bahwa probabilitas untuk semua kelas berjumlah 1), sehingga kelas yang diprediksi akan menjadi 0.

Perancang memiliki penggeser ambang yang berguna untuk meninjau bagaimana performa model akan berubah tergantung pada ambang yang ditetapkan.



1. **Kurva ROC dan metrik AUC**

Istilah lain untuk pengenalan adalah **Tingkat positif benar**, dan ini memiliki metrik yang sesuai bernama **Tingkat positif palsu**, yang mengukur jumlah kasus negatif yang salah diidentifikasi sebagai positif dibandingkan dengan jumlah kasus negatif aktual. Merencanakan metrik ini satu sama lain untuk setiap nilai ambang batas yang mungkin antara 0 dan 1 menghasilkan kurva, yang dikenal sebagai **kurva ROC** (ROC adalah singkatan dari karakteristik operasi penerima, tetapi sebagian besar ilmuwan data hanya menyebutnya kurva ROC). Dalam model yang ideal, kurva akan bergeser ke sisi kiri dan di bagian atas, sehingga mencakup area grafik sepenuhnya. Semakin besar area di bawah kurva, dari metrik **AUC**, (yang dapat berupa nilai apa pun dari 0 hingga 1), semakin baik performa model. Anda dapat meninjau kurva ROC di **Hasil Evaluasi.**



1. **Menyebarkan layanan prediktif**

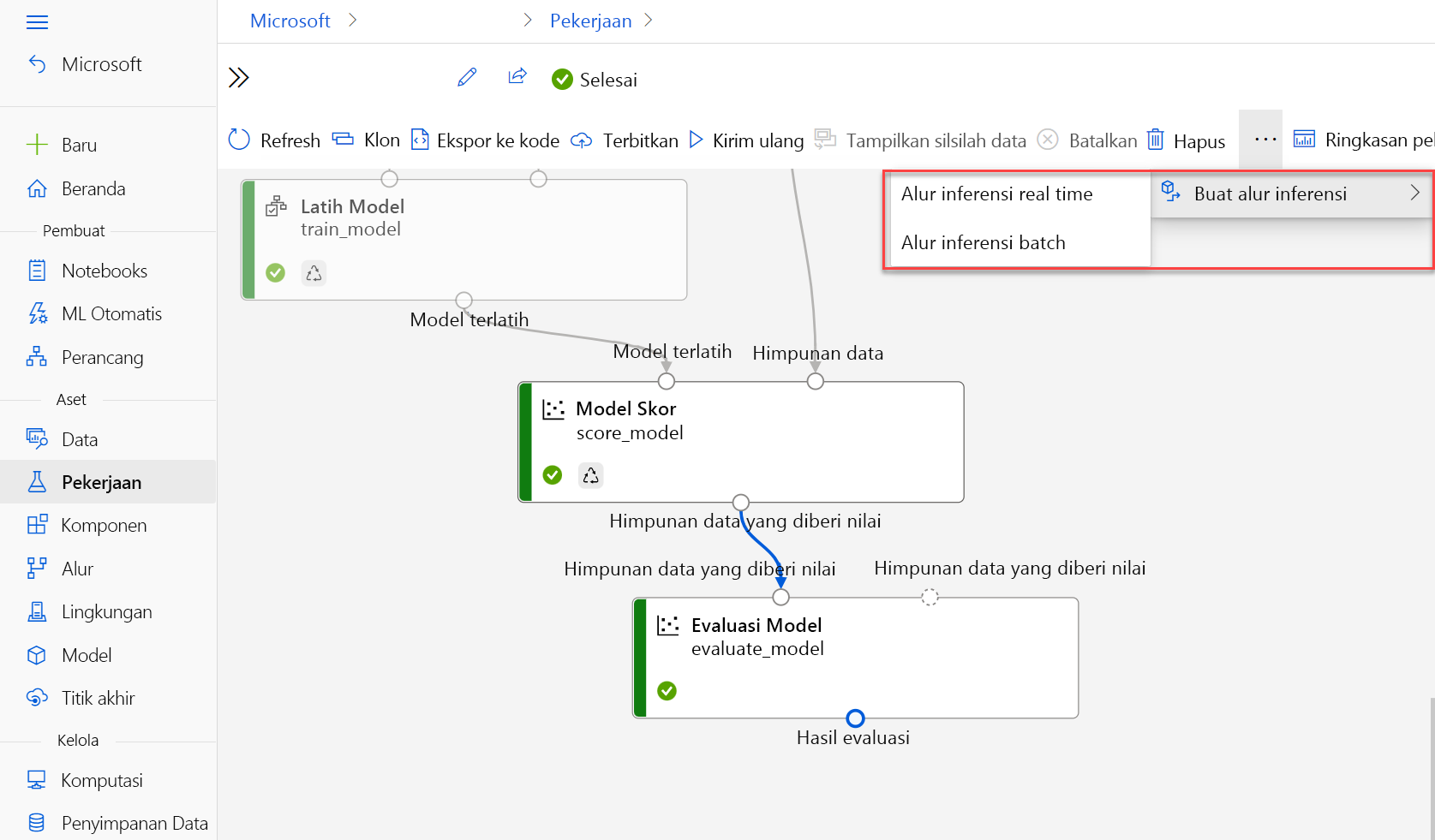
Anda memiliki kemampuan untuk menyebarkan layanan yang dapat digunakan secara real-time. Untuk mengotomatiskan model Anda ke dalam layanan yang membuat prediksi berkelanjutan, Anda perlu membuat dan menyebarkan alur inferensi.

* **Alur inferensi**

Untuk menyebarkan alur, Anda harus terlebih dahulu mengonversi alur pelatihan menjadi alur inferensi real time. Proses ini akan menghapus komponen pelatihan serta menambahkan input dan output layanan web untuk menangani permintaan.

Alur inferensi melakukan transformasi data yang sama seperti alur pertama untuk data baru. Kemudian menggunakan model terlatih untuk menyimpulkan, atau memprediksi, nilai label berdasarkan fitur-fiturnya. Model ini akan menjadi dasar untuk layanan prediktif yang dapat Anda terbitkan untuk digunakan oleh aplikasi.

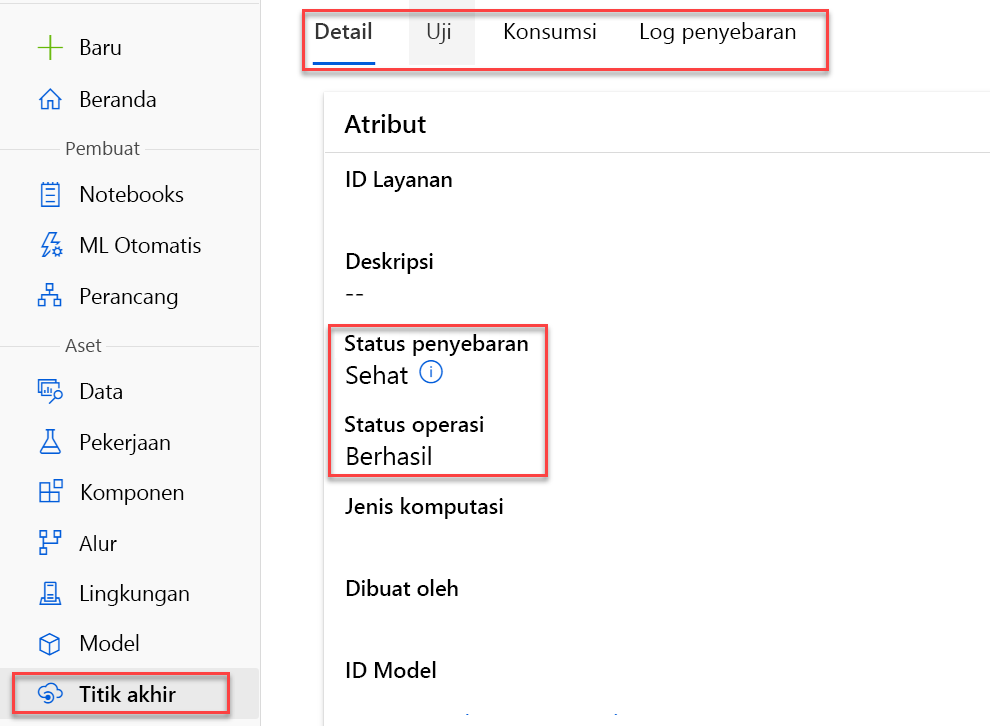
Anda dapat membuat alur inferensi dengan memilih menu di atas pekerjaan yang telah selesai.



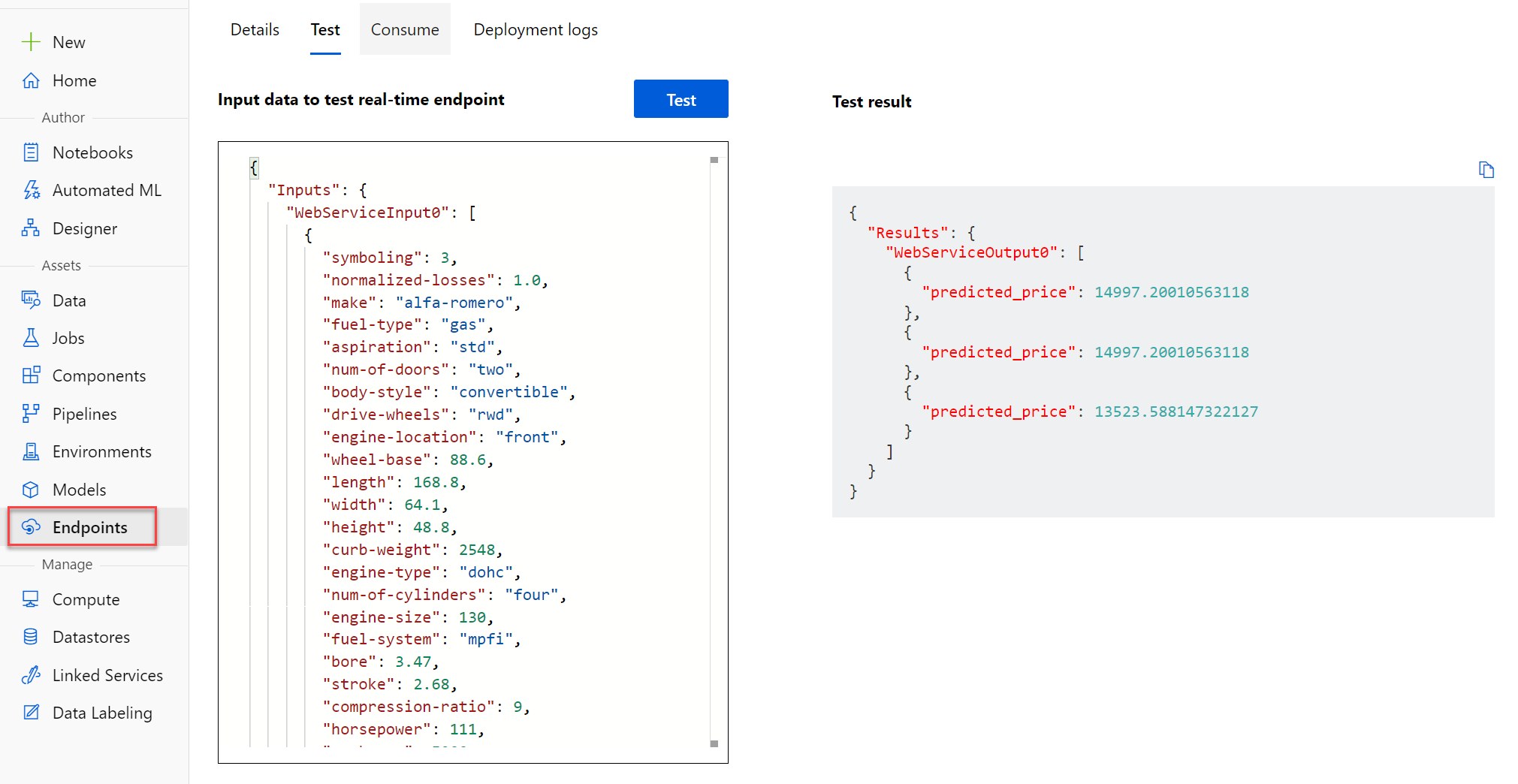
* **Penyebaran**

Setelah membuat alur inferensi, Anda dapat menyebarkannya sebagai titik akhir. Di halaman titik akhir, Anda dapat melihat detail penyebaran, menguji layanan alur Anda dengan data sampel, dan menemukan info masuk untuk menyambungkan layanan alur Anda ke aplikasi klien.

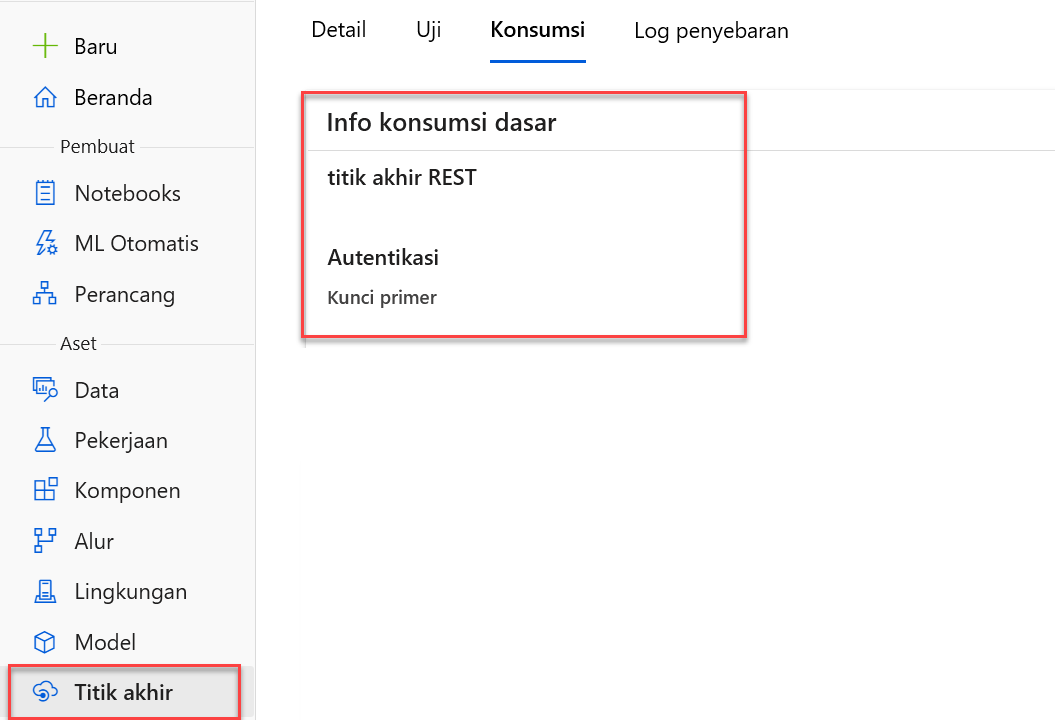
Ini akan memakan waktu beberapa saat agar titik akhir Anda disebarkan. Status Penyebaran pada tab Detail akan menunjukkan keadaan Sehat saat penyebaran berhasil.



Pada tab **Pengujian**, Anda dapat menguji layanan yang disebarkan dengan data sampel dalam format JSON. Tab pengujian adalah alat yang dapat Anda gunakan untuk memeriksa dengan cepat untuk melihat apakah model Anda berperilaku seperti yang diharapkan. Biasanya alat ini sangat membantu untuk menguji layanan sebelum menghubungkannya ke aplikasi.



Anda dapat menemukan info masuk untuk layanan Anda di tab **Pemakaian**. Info masuk ini digunakan untuk menghubungkan model pembelajaran mesin terlatih Anda sebagai layanan ke aplikasi klien.



1. **Latihan - Menjelajahi klasifikasi dengan perancang Azure Machine Learning**

Dalam latihan ini, Anda akan membuat model klasifikasi yang **memprediksi apakah pasien kemungkinan terkena diabetes berdasarkan faktor-faktor seperti BMI, usia, kadar gula, dan indikator lainnya**.

**Catatan** : Untuk menyelesaikan lab ini, Anda memerlukan langganan Azure tempat Anda memiliki akses administratif.

**Link** : <https://microsoftlearning.github.io/AI-900-AIFundamentals/instructions/02b-create-classification-model.html>

1. **Ringkasan**

Dalam modul ini, Anda belajar cara:

* Identifikasi skenario pembelajaran mesin klasifikasi.
* Gunakan perancang Azure Machine Learning untuk melatih model klasifikasi.
* Gunakan model klasifikasi untuk menyimpulkan.
* Menyebarkan dan menguji model klasifikasi.