**Gunakan MLflow untuk melacak eksperimen di Azure Databricks**

Saat menjalankan eksperimen data science dan pembelajaran mesin dalam skala besar, Anda dapat menggunakan MLflow untuk melacak metrik dan jalannya eksperimen.

**Tujuan pembelajaran**

Setelah menyelesaikan modul ini, Anda akan dapat:

* Memahami kemampuan MLflow
* Gunakan terminologi MLflow
* Menjalankan eksperimen

1. **Pengantar**

Saat melakukan tugas pembelajaran mesin di Azure Databricks, Anda dapat menggunakan MLflow untuk melacak dan meninjau pekerjaan Anda. Dalam modul ini, Anda akan mempelajari apa itu MLflow dan bagaimana Anda dapat menggunakan berbagai fiturnya.

1. **Tujuan pembelajaran**

Setelah menyelesaikan modul ini, Anda akan mampu:

* Jelaskan kemampuan MLflow.
* Jelaskan istilah MLflow.
* Mulai lari di MLflow.

1. **Memahami kemampuan MLflow**

**MLflow** adalah produk sumber terbuka yang dirancang untuk mengelola siklus hidup pengembangan Pembelajaran Mesin. Artinya, MLflow memungkinkan ilmuwan data untuk melatih model, mendaftarkan model tersebut, menyebarkan model ke server web, dan mengelola pembaruan model.

1. **Pentingnya MLflow**

MLflow adalah bagian penting dari pembelajaran mesin dengan Azure Databricks, karena MLflow mengintegrasikan proses operasional utama dengan antarmuka Azure Databricks. MLflow memudahkan para ilmuwan data untuk melatih model dan membuatnya tersedia tanpa menulis banyak kode.

Sebagai catatan tambahan, MLflow juga akan beroperasi pada beban kerja di luar Azure Databricks. Contoh dalam modul ini semuanya akan menggunakan Azure Databricks tetapi ini bukan keharusan.

1. **Komponen produk MLflow**

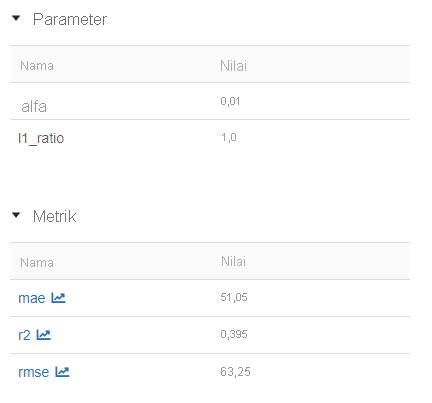
Ada empat komponen untuk MLflow:

* Pelacakan MLflow
* Proyek MLflow
* Model alur ML
* Registri Model MLflow

1. **Pelacakan MLflow**

Pelacakan MLflow memungkinkan ilmuwan data bekerja dengan eksperimen. Untuk setiap proses dalam eksperimen, ilmuwan data dapat mencatat parameter, versi pustaka yang digunakan, metrik evaluasi, dan file keluaran yang dihasilkan saat melatih model pembelajaran mesin.

Pelacakan MLflow memberikan kemampuan untuk mengaudit hasil eksekusi pelatihan model sebelumnya.



1. **Proyek MLflow**

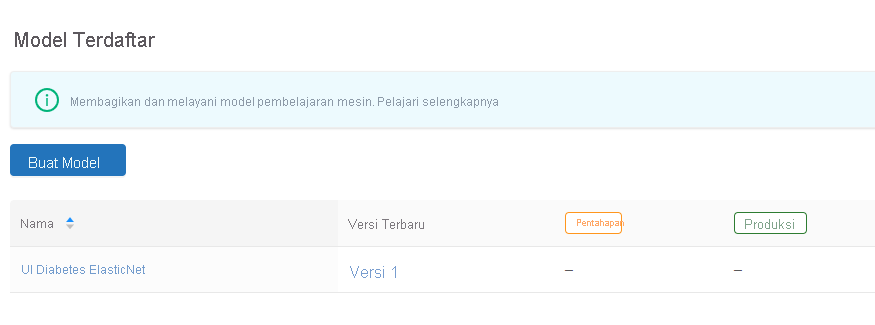
Proyek MLflow adalah cara mengemas kode dengan cara yang memungkinkan penerapan yang konsisten dan kemampuan untuk mereproduksi hasil. MLflow mendukung beberapa lingkungan untuk proyek, termasuk melalui Conda, Docker, dan langsung pada sistem.

1. **Model alur ML**

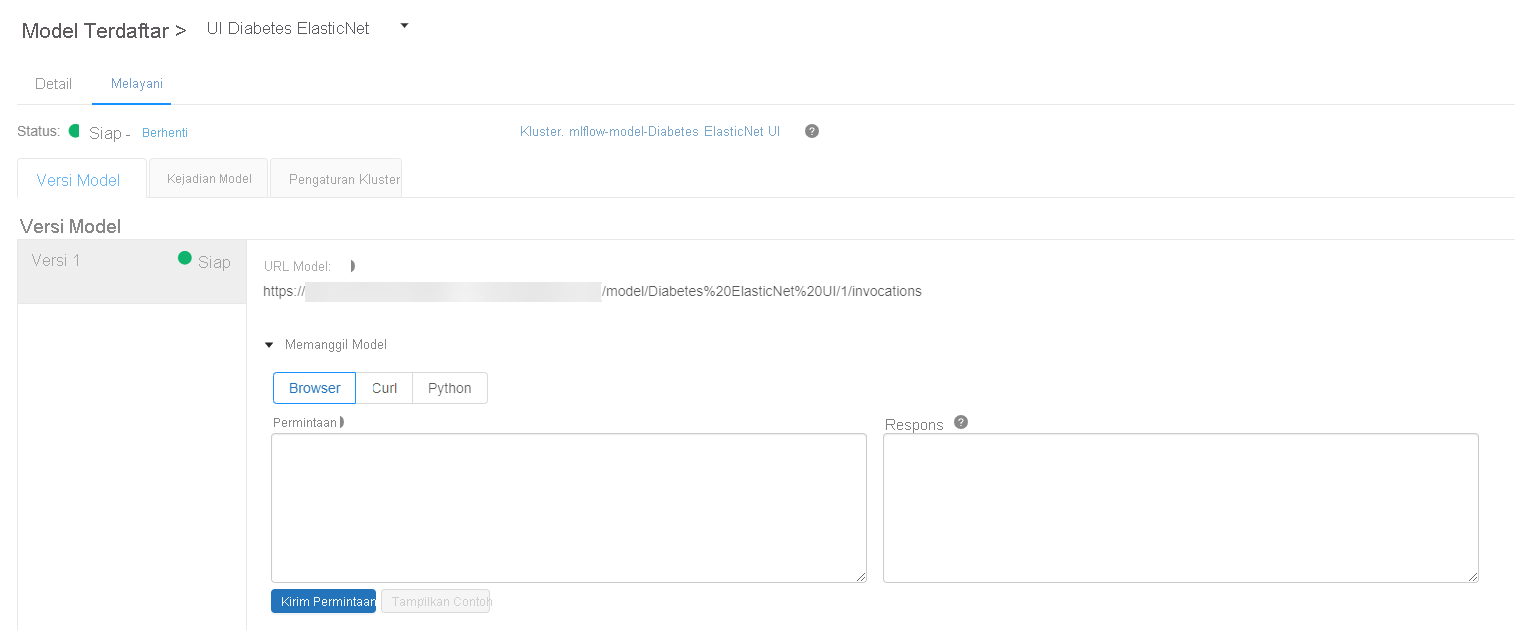
MLflow menawarkan format standar untuk model pengemasan untuk distribusi. Format model standar ini memungkinkan MLflow untuk bekerja dengan model yang dihasilkan dari beberapa pustaka populer, termasuk **scikit-learn**, **Keras**, **MLlib**, **ONNX**, dan banyak lagi. Tinjau **dokumentasi Model MLflow (** <https://mlflow.org/docs/latest/models.html> **)** untuk informasi tentang rangkaian lengkap ragam model yang didukung.

1. **Registri Model MLflow**

Registri Model MLflow memungkinkan ilmuwan data untuk mendaftarkan model dalam registri.



Dari sana, Model MLflow dan Proyek MLflow digabungkan dengan MLflow Model Registry untuk memungkinkan anggota tim operasi menerapkan model dalam registri, melayani mereka baik melalui REST API atau sebagai bagian dari solusi inferensi batch menggunakan Azure Databricks.



1. **Gunakan terminologi MLflow**

Ada beberapa istilah, yang penting untuk dipahami saat bekerja dengan MLflow. Sebagian besar istilah ini cukup umum di ruang ilmu data. Produk lain, seperti Azure Machine Learning, menggunakan terminologi yang sangat mirip untuk memungkinkan pengembangan keterampilan lintas produk yang disederhanakan. Bagian berikut mencakup istilah dan konsep utama untuk setiap produk MLflow.

1. **Pelacakan MLflow**

Pelacakan MLflow dibangun di sekitar **berjalan**, yaitu, eksekusi kode untuk tugas ilmu data. Setiap run berisi beberapa atribut utama, termasuk:

* **Parameter**: Pasangan nilai kunci, yang mewakili masukan. Gunakan parameter untuk melacak **hiperparameter**, yaitu masukan ke fungsi, yang memengaruhi proses pembelajaran mesin.
* **Metrik**: Pasangan nilai kunci, yang menunjukkan performa model. Ini dapat mencakup langkah-langkah evaluasi seperti Root Mean Square Error, dan metrik dapat diperbarui selama proses berjalan. Ini memungkinkan seorang ilmuwan data, misalnya, untuk melacak Root Mean Square Error untuk setiap epoch dari jaringan saraf.
* **Artefak**: File keluaran. Artefak dapat disimpan dalam format apa pun, dan dapat mencakup model, gambar, file log, file data, atau apa pun, yang mungkin penting untuk analisis dan pemahaman model.

Lari ini dapat digabungkan menjadi **eksperimen**, yang dimaksudkan untuk mengumpulkan dan mengatur lari. Misalnya, seorang ilmuwan data dapat membuat eksperimen untuk melatih pengklasifikasi terhadap kumpulan data tertentu. Setiap proses mungkin mencoba algoritma yang berbeda atau set input yang berbeda. Ilmuwan data kemudian dapat meninjau run individu untuk menentukan run mana yang menghasilkan model terbaik.

1. **Proyek MLflow**

**Proyek** di MLflow adalah metode pengemasan kode ilmu data. Ini memungkinkan ilmuwan data lain atau proses otomatis untuk menggunakan kode secara konsisten.

Setiap proyek menyertakan setidaknya satu **titik masuk**, yang merupakan file (baik .**py** atau .**sh**) yang dimaksudkan sebagai titik awal proyek menggunakan. Proyek juga menentukan detail tentang **lingkungan**. Ini termasuk paket khusus (dan versi paket) yang digunakan dalam mengembangkan proyek, karena versi baru dari paket mungkin menyertakan perubahan yang melanggar.

1. **Model alur ML**

**Model** di MLflow adalah direktori yang berisi kumpulan file arbitrer bersama dengan file **MLmodel** di root direktori.

MLflow memungkinkan model memiliki **rasa** tertentu, yang merupakan deskriptor alat atau pustaka mana yang menghasilkan model. Hal ini memungkinkan MLflow untuk bekerja dengan berbagai macam pustaka pemodelan, seperti **scikit-learn, Keras, MLlib, ONNX**, dan banyak lagi. Setiap model memiliki **tanda tangan**, yang menjelaskan masukan dan keluaran yang diharapkan untuk model tersebut.

1. **Registri Model MLflow**

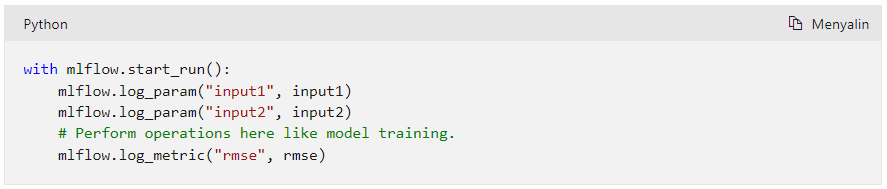
Registri Model MLflow memungkinkan ilmuwan data untuk melacak **model** dari Model MLflow. Dengan kata lain, ilmuwan data **mendaftarkan** model dengan Model Registry, menyimpan detail seperti nama model. Setiap model terdaftar mungkin memiliki beberapa **versi**, yang memungkinkan ilmuwan data melacak perubahan model dari waktu ke waktu.

Dimungkinkan juga untuk **mementaskan** model. Setiap versi model mungkin dalam satu tahap, seperti **Pementasan**, **Produksi**, atau **Diarsipkan**. Ilmuwan dan administrator data dapat **mengalihkan** versi model dari satu tahap ke tahap berikutnya.

1. **Menjalankan eksperimen**

Eksperimen MLflow memungkinkan ilmuwan data melacak pelatihan yang berjalan dalam kumpulan yang disebut **eksperimen**. Ini berguna untuk membandingkan perubahan dari waktu ke waktu atau membandingkan kinerja relatif model dengan nilai hyperparameter yang berbeda.

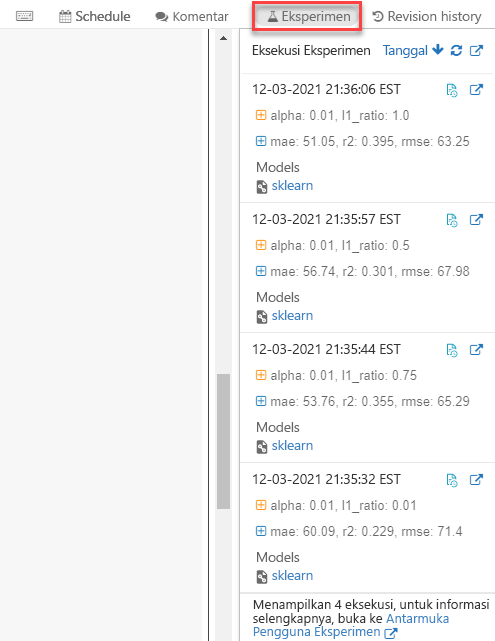
Membuat eksperimen di Azure Databricks terjadi secara otomatis saat Anda memulai lari. Berikut adalah contoh memulai proses di MLflow, mencatat dua parameter, dan mencatat satu metrik:



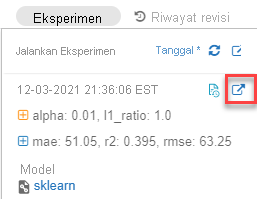
Dalam hal ini, nama eksperimen akan menjadi nama buku catatan. Anda dapat mengekspor variabel bernama **MLFLOW\_EXPERIMENT\_NAME** untuk mengubah nama eksperimen jika Anda memilih.

1. **Meninjau Eksperimen**

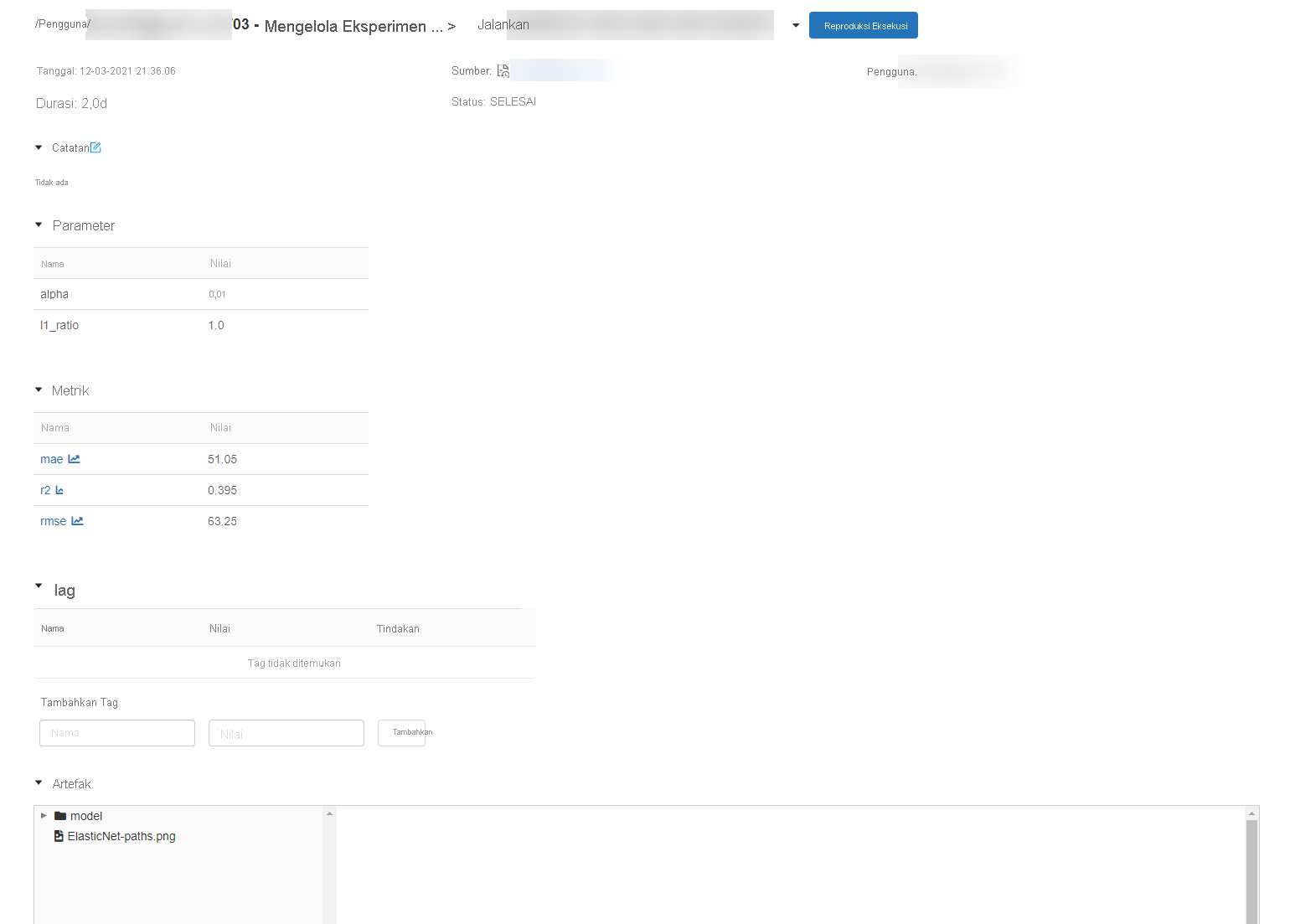
Di dalam buku catatan, opsi menu **Eksperimen** menampilkan bilah konteks, yang mencakup informasi tentang percobaan saat ini.



Memilih ikon Tautan Eksternal dalam percobaan yang dijalankan akan memberikan detail tambahan pada percobaan tertentu.



Tautan ini akan memberikan informasi yang dicatat oleh MLflow Tracker, termasuk catatan, parameter, metrik, tag, dan artefak.



1. **Latihan - Gunakan MLflow untuk melacak eksperimen**

Sekarang, ini adalah kesempatan Anda untuk menggunakan Azure Databricks dan MLflow untuk menjalankan eksperimen dan melacak hasil berbagai pengujian eksperimental.

Dalam latihan ini, Anda akan:

* Menjalankan eksperimen.
* Meninjau metrik eksperimen.

1. **Instruksi**

Ikuti petunjuk berikut untuk menyelesaikan latihan:

1. Buka petunjuk latihan di <https://aka.ms/mslearn-dp090.> \ <https://microsoftlearning.github.io/dp-090-databricks-ml/>
2. Selesaikan latihan **Menggunakan MLflow untuk Melacak Eksperimen**.
3. **Ringkasan**

Dalam modul ini, Anda mempelajari cara mengelola eksperimen menggunakan MLflow di Azure Databricks.

Sekarang setelah Anda menyelesaikan modul ini, Anda dapat:

* Jelaskan kemampuan MLflow.
* Jelaskan istilah MLflow.
* Mulai lari di MLflow.