



AWS Panduan Keputusan

Memilih layanan pembelajaran AWS mesin



Memilih layanan pembelajaran AWS mesin: AWS Panduan Keputusan

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Merek dagang dan tampilan dagang Amazon tidak dapat digunakan sehubungan dengan produk atau layanan yang bukan milik Amazon, dalam bentuk apa pun yang mungkin menimbulkan kebingungan di kalangan pelanggan, atau dalam bentuk apa pun yang merendahkan atau mendiskreditkan Amazon. Semua merek dagang lain yang tidak dimiliki oleh Amazon adalah properti dari pemiliknya masing-masing, yang mungkin atau mungkin tidak berafiliasi dengan, berhubungan dengan, atau disponsori oleh Amazon.

Table of Contents

Panduan Keputusan	1
Pengantar	1
Memahami	2
Pertimbangkan	4
Pilih	7
Gunakan	10
Jelajahi	22
Sumber daya	23
Riwayat dokumen	25
.....	xxvi

Memilih layanan pembelajaran AWS mesin

Pilih layanan dan kerangka kerja yang tepat untuk mendukung pekerjaan Anda

Tujuan	Bantu menentukan layanan AWS ML mana yang paling sesuai dengan kebutuhan Anda.
Terakhir diperbarui	3 Mei 2024
Layanan yang tercakup	• AI Augmented AI Amazon • Amazon CodeGuru • Amazon Comprehend • DevOpsGuru Amazon • Amazon Forecast • Amazon Kendra • Amazon Lex • Amazon Personalisasi • Amazon Polly • Amazon Rekognition • Amazon SageMaker AI • Amazon Texttract • Amazon Transcribe • Amazon Translate

Pengantar

Pada dasarnya, pembelajaran mesin (ML) dirancang untuk menyediakan alat dan layanan digital untuk belajar dari data, mengidentifikasi pola, membuat prediksi, dan kemudian bertindak berdasarkan prediksi tersebut. Hampir semua sistem kecerdasan buatan (AI) saat ini dibuat menggunakan ML. ML menggunakan sejumlah besar data untuk membuat dan memvalidasi logika keputusan. Logika keputusan ini membentuk dasar dari model AI.

Skenario di mana layanan pembelajaran AWS mesin dapat diterapkan meliputi:

- Kasus penggunaan khusus — layanan pembelajaran AWS mesin dapat mendukung kasus penggunaan bertenaga AI Anda dengan berbagai algoritme, model, dan solusi pra-bangun untuk kasus penggunaan umum dan industri. Anda memiliki pilihan 23 layanan pra-terlatih, termasuk Amazon Personalize, Amazon Kendra, dan Amazon Monitron.
- Menyesuaikan dan menskalakan machine learning — Amazon SageMaker AI dirancang untuk membantu Anda membangun, melatih, dan menerapkan model ML untuk kasus penggunaan apa pun. Anda dapat membuat sendiri atau mengakses model dasar open source AWS melalui Amazon SageMaker AI dan Amazon Bedrock.
- Mengakses infrastruktur khusus - Gunakan kerangka kerja dan infrastruktur ML yang disediakan AWS saat Anda memerlukan fleksibilitas dan kontrol yang lebih besar atas alur kerja pembelajaran mesin Anda, dan bersedia mengelola sendiri infrastruktur dan sumber daya yang mendasarinya.

Panduan keputusan ini akan membantu Anda mengajukan pertanyaan yang tepat, mengevaluasi kriteria dan masalah bisnis Anda, dan menentukan layanan mana yang paling sesuai dengan kebutuhan Anda.

Memahami

Ketika organisasi terus mengadopsi teknologi AI dan ML, pentingnya memahami dan memilih di antara layanan AWS ML merupakan tantangan yang sedang berlangsung.

AWS menyediakan berbagai layanan ML yang dirancang untuk membantu organisasi membangun, melatih, dan menerapkan model ML dengan lebih cepat dan mudah. Layanan ini dapat digunakan untuk memecahkan berbagai masalah bisnis seperti prediksi churn pelanggan, deteksi penipuan, dan pengenalan gambar dan ucapan.

What is it?



Artificial intelligence (AI)

Any technique that enables computers to mimic human intelligence using logic, if-then statements, and machine learning



Machine learning (ML)

A subset of AI that uses machines to search for patterns in data to build logic models automatically



Classification AI and Predictive AI

A subset of ML that recognizes patterns to identify something (Classification AI) or predicts future trends based on statistical patterns and historical data (Predictive AI)



Generative AI

A subset of DL that can create new content and ideas powered by large, pretrained models called foundation models (FMs)

Sebelum menyelam lebih dalam ke layanan AWS ML, mari kita lihat hubungan antara AI dan ML.

- Pada tingkat tinggi, kecerdasan buatan adalah cara untuk menggambarkan sistem apa pun yang dapat mereplikasi tugas yang sebelumnya membutuhkan kecerdasan manusia. Sebagian besar kasus penggunaan AI mencari hasil probabilistik—membuat prediksi atau keputusan dengan tingkat kepastian yang tinggi, mirip dengan penilaian manusia.
- Hampir semua sistem AI saat ini dibuat menggunakan pembelajaran mesin. ML menggunakan sejumlah besar data untuk membuat dan memvalidasi logika keputusan, yang dikenal sebagai model.
- Klasifikasi AI adalah bagian dari ML yang mengenali pola untuk mengidentifikasi sesuatu. Predictive AI adalah bagian dari ML yang memprediksi tren masa depan berdasarkan pola statistik data historis.
- Akhirnya, AI generatif adalah bagian dari pembelajaran mendalam yang dapat membuat konten dan ide baru, seperti percakapan, cerita, gambar, video, dan musik. Generative AI didukung oleh model yang sangat besar yang telah dilatih sebelumnya pada kumpulan data yang luas, yang disebut Foundation Models atau FMs [Amazon Bedrock](#) adalah layanan yang dikelola sepenuhnya yang menawarkan pilihan kinerja tinggi FMs untuk membangun dan menskalakan aplikasi AI

generatif. [Amazon Q Developer](#) dan [Amazon Q Business](#) adalah asisten bertenaga AI generasi untuk kasus penggunaan tertentu.

Panduan ini dirancang terutama untuk mencakup layanan dalam kategori pembelajaran mesin AI Klasifikasi dan Prediktif AI.

Selain itu, AWS menawarkan perangkat keras khusus yang dipercepat untuk pelatihan dan inferensi MLM berkinerja tinggi.

- Instans [Amazon EC2 P5](#) dilengkapi dengan NVIDIA H100 Tensor Core GPUs, yang sangat cocok untuk tugas pelatihan dan inferensi dalam pembelajaran mesin. Instans [Amazon EC2 G5](#) menampilkan hingga 8 NVIDIA A10G Tensor Core GPUs, dan prosesor AMD EPYC generasi kedua, untuk berbagai kasus penggunaan intensif grafis dan pembelajaran mesin.
- [AWS Trainium](#) adalah akselerator ML generasi kedua yang AWS memiliki tujuan untuk pelatihan pembelajaran mendalam (DL) model parameter 100B+.
- [AWS Inferentia Instans Amazon EC2 Inf2 berbasis](#) 2 dirancang untuk memberikan kinerja tinggi dengan biaya terendah di Amazon EC2 untuk aplikasi inferensi DL dan AI generatif Anda.

Pertimbangkan

Saat memecahkan masalah bisnis dengan layanan AWS ML, pertimbangan beberapa kriteria utama dapat membantu memastikan kesuksesan. Bagian berikut menguraikan beberapa kriteria utama yang perlu dipertimbangkan saat memilih layanan ML.

Problem definition

Definisi masalah

Langkah pertama dalam siklus hidup ML adalah membingkai masalah bisnis. Memahami masalah yang Anda coba pecahkan sangat penting untuk memilih layanan AWS ML yang tepat, karena layanan yang berbeda dirancang untuk mengatasi masalah yang berbeda. Penting juga untuk menentukan apakah ML paling cocok untuk masalah bisnis Anda.

Setelah Anda menentukan bahwa ML adalah yang paling cocok, Anda dapat mulai dengan memilih dari berbagai layanan AWS AI yang dibuat khusus (di berbagai bidang seperti ucapan, visi, dan dokumen).

Amazon SageMaker AI menyediakan infrastruktur yang dikelola sepenuhnya jika Anda perlu membangun dan melatih model Anda sendiri. AWS menawarkan serangkaian kerangka kerja dan pilihan infrastruktur yang canggih untuk kasus di mana Anda memerlukan model ML yang sangat disesuaikan dan terspesialisasi. AWS juga menawarkan serangkaian model fondasi populer untuk membangun aplikasi baru dengan AI generatif.

ML algorithm

Algoritma ML

Memilih algoritma ML untuk masalah bisnis yang Anda coba selesaikan tergantung pada jenis data yang Anda kerjakan, serta hasil yang diinginkan. Informasi berikut menguraikan bagaimana masing-masing kategori layanan AWS AI/ML utama memberdayakan Anda untuk bekerja dengan algoritmanya:

- **Layanan AI khusus:** Layanan ini menawarkan kemampuan terbatas untuk menyesuaikan algoritme ML, karena merupakan model pra-pelatihan yang dioptimalkan untuk tugas-tugas tertentu. Anda biasanya dapat menyesuaikan data input dan beberapa parameter, tetapi tidak memiliki akses ke model ML yang mendasarinya atau kemampuan untuk membuat model Anda sendiri.
- **Amazon SageMaker AI:** Layanan ini memberikan fleksibilitas dan kontrol paling besar atas algoritma ML. Anda dapat menggunakan SageMaker AI untuk membuat model khusus menggunakan algoritme dan kerangka kerja Anda sendiri, atau menggunakan model dan algoritme pra-bangun yang disediakan oleh AWS. Hal ini memungkinkan untuk tingkat tinggi kustomisasi dan kontrol atas proses ML.
- **Kerangka kerja dan infrastruktur ML tingkat rendah:** Layanan ini menawarkan fleksibilitas dan kontrol paling besar atas algoritme ML. Anda dapat menggunakan layanan ini untuk membuat model ML yang sangat disesuaikan menggunakan algoritme dan kerangka kerjanya sendiri. Namun, menggunakan layanan ini membutuhkan keahlian ML yang signifikan dan mungkin tidak layak untuk semua kasus penggunaan.

Security

Keamanan

Jika Anda memerlukan titik akhir pribadi di VPC Anda, opsi Anda akan bervariasi berdasarkan lapisan layanan ML AWS yang Anda gunakan. Ini termasuk:

- Layanan AI khusus: Sebagian besar layanan AI khusus saat ini tidak mendukung titik akhir pribadi di VPCs. Namun, Amazon Rekognition Custom Labels dan Amazon Comprehend Custom dapat diakses menggunakan endpoint VPC.
- Layanan AI inti: Amazon Translate, Amazon Transcribe, dan Amazon Comprehend semua mendukung titik akhir VPC.
- Amazon SageMaker AI: SageMaker AI menyediakan dukungan bawaan untuk titik akhir VPC, memungkinkan Anda untuk menerapkan model terlatih mereka sebagai titik akhir yang hanya dapat diakses dari dalam VPC mereka.
- Kerangka kerja dan infrastruktur ML tingkat rendah: Anda dapat menerapkan model Anda di EC2 instans Amazon atau dalam kontainer dalam VPC Anda, memberikan kontrol penuh atas konfigurasi jaringan.

Latency

Latensi

Layanan AI tingkat tinggi, seperti Amazon Rekognition dan Amazon Transcribe, dirancang untuk menangani berbagai macam kasus penggunaan dan menawarkan kinerja tinggi dalam hal kecepatan. Namun, mereka mungkin tidak memenuhi persyaratan latensi tertentu.

Jika Anda menggunakan kerangka kerja dan infrastruktur ML tingkat rendah, kami sarankan untuk memanfaatkan Amazon AI. SageMaker Opsi ini umumnya lebih cepat daripada membangun model khusus karena layanan yang dikelola sepenuhnya dan opsi penerapan yang dioptimalkan. Sementara model kustom yang sangat dioptimalkan mungkin SageMaker mengungguli AI, itu akan membutuhkan keahlian dan sumber daya yang signifikan untuk membangunnya.

Accuracy

Akurasi

Keakuratan layanan AWS ML bervariasi berdasarkan kasus penggunaan spesifik dan tingkat kustomisasi yang diperlukan. Layanan AI tingkat tinggi, seperti Amazon Rekognition, dibangun di atas model pra-terlatih yang telah dioptimalkan untuk tugas-tugas tertentu dan menawarkan akurasi tinggi dalam banyak kasus penggunaan.

Dalam beberapa kasus, Anda dapat memilih untuk menggunakan Amazon SageMaker AI, yang menyediakan platform yang lebih fleksibel dan dapat disesuaikan untuk membangun dan melatih model HTML khusus. Dengan membangun model Anda sendiri, Anda mungkin dapat mencapai

akurasi yang lebih tinggi daripada yang mungkin dilakukan dengan model yang telah dilatih sebelumnya.

Anda juga dapat memilih untuk menggunakan kerangka kerja dan infrastruktur HTML, seperti TensorFlow dan Apache MXNet, untuk membangun model yang sangat disesuaikan yang menawarkan akurasi setinggi mungkin untuk kasus penggunaan spesifik Anda.

AWS and responsible AI

AWS dan AI yang bertanggung jawab

AWS membangun model pondasi (FMs) dengan AI yang bertanggung jawab dalam pikiran pada setiap tahap proses pengembangannya. Sepanjang desain, pengembangan, penyebaran, dan operasi, kami mempertimbangkan berbagai faktor termasuk:

1. Akurasi (seberapa dekat ringkasan cocok dengan dokumen yang mendasarinya; apakah biografi benar secara faktual)
2. Keadilan, (apakah output memperlakukan kelompok demografis dengan cara yang sama)
3. Pertimbangan kekayaan intelektual dan hak cipta
4. Penggunaan yang tepat (menyaring permintaan pengguna untuk nasihat hukum, atau diagnosa medis, atau kegiatan ilegal)
5. Toksisitas (ujaran kebencian, kata-kata kotor, dan penghinaan)
6. Privasi (melindungi informasi pribadi dan permintaan pelanggan)

AWS membangun solusi untuk mengatasi masalah ini ke dalam proses yang digunakan untuk memperoleh data pelatihan, ke dalam FMs diri mereka sendiri, dan ke dalam teknologi yang digunakan untuk pra-proses permintaan pengguna dan output pasca-proses.

Pilih

Sekarang setelah Anda mengetahui kriteria yang akan digunakan untuk mengevaluasi opsi layanan ML Anda, Anda siap untuk memilih layanan AWS ML mana yang tepat untuk kebutuhan organisasi Anda. Tabel berikut menyoroti layanan ML mana yang dioptimalkan untuk keadaan apa. Gunakan untuk membantu menentukan layanan AWS ML yang paling cocok untuk kasus penggunaan Anda.

Kategori	Kapan Anda akan menggunakannya?	Untuk apa itu dioptimalkan?	AI/ML Layanan atau lingkungan terkait
Kasus penggunaan khusus Layanan kecerdasan buatan ini dimaksudkan untuk memenuhi kebutuhan spesifik. Mereka termasuk personalisasi, peramalan, deteksi anomali, transkripsi ucapan, dan lain-lain. Karena mereka dikirimkan sebagai layanan, mereka dapat disematkan ke dalam aplikasi tanpa memerlukan keahlian ML apa pun.	Gunakan layanan AI yang disediakan AWS saat Anda memerlukan fungsionalitas khusus yang dibuat sebelumnya untuk diintegrasikan ke dalam aplikasi Anda, tanpa perlu penyesuaian ekstensif atau keahlian pembelajaran mesin. Layanan ini dirancang agar mudah digunakan dan tidak memerlukan banyak pengkodean atau konfigurasi.	Layanan ini dirancang agar mudah digunakan dan tidak memerlukan banyak pengkodean, konfigurasi, atau keahlian ML.	AI Augmented AI Amazon Amazon CodeGuru Amazon Comprehend Amazon Comprehend Medical DevOpsGuru Amazon Amazon Forecast Amazon Kendra Amazon Lex Amazon Personalisasi Amazon Polly Amazon Rekognition Amazon Texttract Amazon Transcribe Amazon Translate
Layanan ML Layanan ini dapat digunakan untuk mengembangkan model pembelajaran mesin yang disesuaikan atau	Gunakan layanan ini ketika Anda membutuhkan model pembelajaran mesin atau alur kerja yang lebih disesuaikan yang melampaui fungsionalitas	Layanan ini dioptimalkan untuk membangun dan melatih model pembelajaran mesin khusus, pelatihan skala besar pada beberapa instance	Amazon SageMaker AI Amazon SageMaker AI JumpStart SageMaker Studio AI SageMaker Kanvas AI

Kategori	Kapan Anda akan menggunakannya?	Untuk apa itu dioptimalkan?	AI/ML Layanan atau lingkungan terkait
alur kerja yang melampaui fungsionalitas pra-bangun yang ditawarkan oleh layanan AI inti.	litas pra-bangun yang ditawarkan oleh layanan AI inti.	atau cluster GPU, kontrol lebih besar atas penerapan model pembelajaran mesin, inferensi waktu nyata, dan untuk membangun alur kerja. end-to-end	SageMaker Lab Studio AI SageMaker AI Ground Truth PyTorch pada AWS Apache MxNet Hugging Face TensorFlow pada AWS
Infrastruktur Untuk menerapkan pembelajaran mesin dalam produksi, Anda memerlukan infrastruktur hemat biaya, yang diaktifkan Amazon dengan AWS silikon buatan.	Gunakan saat Anda ingin mencapai biaya terendah untuk model pelatihan dan perlu menjalankan inferensi di cloud.	Dioptimalkan untuk mendukung penerapan pembelajaran mesin yang hemat biaya.	AWS Trainium AWS Inferentia dan Inferentia2 Amazon SageMaker AI HyperPod
Alat dan layanan terkait Alat dan layanan terkait ini dirancang untuk membantu Anda memudahkan penerapan pembelajaran mesin.	Layanan dan alat ini dirancang untuk membantu Anda mempercepat pembelajaran mendalam di cloud, menyediakan gambar mesin Amazon, gambar docker, dan resolusi entitas.	Dioptimalkan untuk membantu Anda mempercepat pembelajaran mendalam di cloud.	AWS Deep Learning AMIs AWS Deep Learning Containers AWS Resolusi Entitas

Gunakan

Sekarang setelah Anda memiliki pemahaman yang jelas tentang kriteria yang perlu Anda terapkan dalam memilih layanan AWS ML, Anda dapat memilih layanan AWS AI/ML mana yang dioptimalkan untuk kebutuhan bisnis Anda.

Untuk mengeksplorasi cara menggunakan dan mempelajari lebih lanjut tentang layanan yang telah Anda pilih, kami telah menyediakan tiga set jalur untuk mengeksplorasi cara kerja setiap layanan. Kumpulan jalur pertama menyediakan dokumentasi mendalam, tutorial langsung, dan sumber daya untuk memulai Amazon Comprehend, Amazon Texttract, Amazon Translate, Amazon Lex, Amazon Polly, Amazon Rekognition, dan Amazon Transcribe.

Amazon Comprehend

- Memulai Amazon Comprehend

Gunakan konsol Amazon Comprehend untuk membuat dan menjalankan tugas deteksi entitas asinkron.

[Mulai dengan tutorial”](#)

- Analisis wawasan dalam teks dengan Amazon Comprehend

Pelajari cara menggunakan Amazon Comprehend untuk menganalisis dan memperoleh wawasan dari teks.

[Mulai dengan tutorial”](#)

- Amazon Comprehend Harga

Jelajahi informasi tentang Amazon Comprehend harga dan contoh.

[Jelajahi panduannya”](#)

Amazon Texttract

- Memulai dengan Amazon Texttract

Pelajari bagaimana Amazon Textract dapat digunakan dengan teks yang diformat untuk mendeteksi kata dan baris kata yang terletak berdekatan satu sama lain, serta menganalisis dokumen untuk item seperti teks terkait, tabel, pasangan nilai kunci, dan elemen pilihan.

[Jelajahi panduannya](#)

- Ekstrak teks dan data terstruktur dengan Amazon Textract

Pelajari cara menggunakan Amazon Textract untuk mengekstrak teks dan data terstruktur dari dokumen.

[Mulai dengan tutorial](#)

- AWS Power Hour: Machine Learning

Selami Amazon Textract di episode ini, habiskan waktu di dalamnya AWS Management Console, dan tinjau contoh kode yang akan membantu Anda memahami cara memanfaatkan layanan sebaik mungkin. APIs

[Tonton videonya](#)

Amazon Translate

- Memulai Amazon Translate menggunakan konsol

Cara termudah untuk memulai Amazon Translate adalah dengan menggunakan konsol untuk menerjemahkan beberapa teks. Pelajari cara menerjemahkan hingga 10.000 karakter menggunakan konsol.

[Jelajahi panduannya](#)

- Terjemahkan Teks Antar Bahasa di Cloud

Dalam contoh tutorial ini, sebagai bagian dari perusahaan manufaktur bagasi internasional, Anda perlu memahami apa yang dikatakan pelanggan tentang produk Anda dalam ulasan dalam bahasa pasar lokal - Prancis.

[Mulai dengan tutorial](#)

- Harga Amazon Translate

Jelajahi harga Amazon Translate, termasuk Tingkat Gratis - yang menyediakan 2 juta karakter per bulan selama 12 bulan.

[Jelajahi panduannya](#)

Amazon Lex

- Panduan Pengembang Amazon Lex V2

Jelajahi informasi tentang memulai, cara kerjanya, dan informasi harga untuk Amazon Lex V2.

[Jelajahi panduannya](#)

- Pengantar Amazon Lex Kami memperkenalkan Anda ke layanan percakapan Amazon Lex, dan memandu Anda melalui contoh yang menunjukkan cara membuat bot dan menyebarkannya ke berbagai layanan obrolan.

[Ikuti kursus](#)” (diperlukan masuk)

- Menjelajahi AI Generatif dalam pengalaman percakapan

Jelajahi penggunaan AI generatif dalam pengalaman percakapan.

[Baca blog](#)”

Amazon Polly

- Apa itu Amazon Polly?

Jelajahi ikhtisar lengkap layanan cloud yang mengubah teks menjadi ucapan yang hidup, dan dapat digunakan untuk mengembangkan aplikasi untuk meningkatkan keterlibatan dan aksesibilitas pelanggan Anda.

[Jelajahi panduannya](#)

- Sorot teks saat diucapkan menggunakan Amazon Polly

Kami memperkenalkan Anda pada pendekatan untuk menyoroti teks saat sedang diucapkan untuk menambahkan kemampuan visual ke audio di buku, situs web, blog, dan pengalaman digital lainnya.

[Baca blog](#)

- Buat audio untuk konten dalam berbagai bahasa dengan persona suara TTS yang sama di Amazon Polly

Kami menjelaskan Neural Text-to-Speech (NTTS) dan mendiskusikan bagaimana portofolio luas suara yang tersedia, menyediakan berbagai penutur berbeda dalam bahasa yang didukung, dapat bekerja untuk Anda.

[Baca blog](#)

Amazon Rekognition

- Apa itu Amazon Rekognition?

Jelajahi bagaimana Anda dapat menggunakan layanan ini untuk menambahkan analisis gambar dan video ke aplikasi Anda.

[Jelajahi panduannya](#)

- Rekognition Langsung: Analisis Gambar dan Video Otomatis

Pelajari cara kerja pengenalan wajah dengan streaming video, bersama dengan contoh kode dan poin-poin penting dengan kecepatan yang dipandu sendiri.

[Mulai dengan tutorial](#)

- Amazon Rekognition FAQs

Pelajari dasar-dasar Amazon Rekognition dan bagaimana hal itu dapat membantu Anda meningkatkan pembelajaran mendalam dan menganalisis aplikasi Anda secara visual.

[Baca FAQs](#)

Amazon Transcribe

- Apa itu Amazon Transcribe?

Jelajahi layanan pengenalan suara AWS otomatis menggunakan ML untuk mengonversi audio menjadi teks. Pelajari cara menggunakan layanan ini sebagai transkripsi mandiri atau menambahkan speech-to-text kemampuan ke aplikasi apa pun.

[Jelajahi panduannya](#)

- Harga Amazon Transcribe

Kami memperkenalkan AWS pay-as-you-go transkripsi kepada Anda, termasuk opsi model bahasa khusus dan Tingkat Gratis Amazon Transcribe.

[Jelajahi panduannya](#)

- Buat transkrip audio dengan Amazon Transcribe

Pelajari cara menggunakan Amazon Transcribe untuk membuat transkrip teks file audio yang direkam menggunakan skenario kasus penggunaan dunia nyata untuk menguji kebutuhan Anda.

[Mulai dengan tutorial](#)

- Membangun aplikasi streaming Amazon Transcribe

Pelajari cara membuat aplikasi untuk merekam, mentranskripsikan, dan menerjemahkan audio langsung secara real-time, dengan hasil yang dikirimkan langsung kepada Anda melalui email.

[Jelajahi panduannya](#)

Rangkaian Layanan AWS jalur AI/ML kedua menyediakan dokumentasi mendalam, tutorial langsung, dan sumber daya untuk memulai layanan di keluarga Amazon AI. SageMaker

SageMaker AI

- Bagaimana Amazon SageMaker AI bekerja

[Jelajahi ikhtisar pembelajaran mesin dan cara kerja SageMaker AI.](#)

[Jelajahi panduannya”](#)

- Memulai dengan Amazon SageMaker AI

Pelajari cara bergabung dengan Domain SageMaker AI Amazon, memberi Anda akses ke Amazon SageMaker AI Studio dan RStudio SageMaker AI.

[Jelajahi panduannya”](#)

- Gunakan Apache Spark dengan Amazon AI SageMaker

Pelajari cara menggunakan Apache Spark untuk preprocessing data dan SageMaker AI untuk pelatihan model dan hosting.

[Jelajahi panduannya”](#)

- Gunakan kontainer Docker untuk membuat model

Jelajahi bagaimana Amazon SageMaker AI memanfaatkan kontainer Docker secara ekstensif untuk tugas build dan runtime. Pelajari cara menerapkan image Docker yang sudah dibuat sebelumnya untuk algoritme bawaannya dan kerangka kerja pembelajaran mendalam yang didukung yang digunakan untuk pelatihan dan inferensi.

[Jelajahi panduannya”](#)

- Kerangka kerja dan bahasa pembelajaran mesin

Pelajari cara memulai SageMaker AI menggunakan Amazon SageMaker AI Python SDK.

[Jelajahi panduannya”](#)

SageMaker AI Autopilot

- Buat eksperimen Amazon SageMaker AI Autopilot untuk data tabular

Pelajari cara membuat eksperimen Amazon SageMaker AI Autopilot untuk mengeksplorasi, melakukan pra-proses, dan melatih berbagai kandidat model pada kumpulan data tabular.

[Jelajahi panduannya”](#)

- Secara otomatis membuat model pembelajaran mesin

Pelajari cara menggunakan Amazon SageMaker AI Autopilot untuk membuat, melatih, dan menyetel model ML secara otomatis, serta menerapkan model untuk membuat prediksi.

[Mulai dengan tutorial](#)

- Jelajahi pemodelan dengan Amazon SageMaker AI Autopilot dengan contoh notebook ini

Jelajahi contoh notebook untuk pemasaran langsung, prediksi churn pelanggan, dan cara membawa kode pemrosesan data Anda sendiri ke Amazon SageMaker AI Autopilot.

[Jelajahi panduannya](#)

SageMaker AI Canvas

- Mulai menggunakan Amazon SageMaker AI Canvas

Pelajari cara memulai menggunakan SageMaker AI Canvas.

[Jelajahi panduannya](#)

- Hasilkan prediksi pembelajaran mesin tanpa menulis kode

Tutorial ini menjelaskan cara menggunakan Amazon SageMaker AI Canvas untuk membangun model ML dan menghasilkan prediksi yang akurat tanpa menulis satu baris kode pun.

[Mulai dengan tutorial](#)

- Selami lebih dalam ke SageMaker AI Canvas

Jelajahi tampilan mendalam pada SageMaker AI Canvas dan kemampuan visualnya, tanpa kode HTML.

[Baca blog](#)

- Gunakan Amazon SageMaker AI Canvas untuk membuat Model ML pertama Anda

Pelajari cara menggunakan Amazon SageMaker AI Canvas untuk membuat model ML guna menilai retensi pelanggan, berdasarkan kampanye email untuk produk dan layanan baru.

[Memulai dengan lab](#)

SageMaker AI Data Wrangler

- Memulai dengan Amazon SageMaker AI Data Wrangler

Jelajahi cara mengatur SageMaker AI Data Wrangler dan kemudian berikan panduan menggunakan kumpulan data contoh yang ada.

[Jelajahi panduannya](#)

- Siapkan data pelatihan untuk pembelajaran mesin dengan kode minimal

Pelajari cara menyiapkan data untuk ML menggunakan Amazon SageMaker AI Data Wrangler.

[Mulai dengan tutorial](#)

- SageMaker Lokakarya menyelam mendalam AI Data Wrangler

Pelajari cara menerapkan jenis analisis yang sesuai pada kumpulan data Anda untuk mendeteksi anomali dan masalah, gunakan turunan results/insights untuk merumuskan tindakan perbaikan selama transformasi pada kumpulan data Anda, dan uji pilihan dan urutan transformasi yang tepat menggunakan opsi pemodelan cepat yang disediakan oleh AI Data Wrangler. SageMaker

[Mulai dengan Workshop](#)

SageMaker AI Ground Truth

- Memulai dengan Amazon Ground Truth

Jelajahi cara menggunakan konsol untuk membuat pekerjaan pelabelan, menetapkan tenaga kerja publik atau swasta, dan mengirim pekerjaan pelabelan ke tenaga kerja Anda. Pelajari cara memantau kemajuan pekerjaan pelabelan.

[Jelajahi panduannya”](#)

- Data Pelatihan Label untuk Machine Learning

Pelajari cara menyiapkan pekerjaan pelabelan di Amazon SageMaker AI Ground Truth untuk membuat anotasi data pelatihan untuk model ML Anda.

[Mulai dengan tutorial”](#)

- Memulai dengan Amazon Ground Truth Plus Jelajahi cara menyelesaikan langkah-langkah yang diperlukan untuk memulai proyek Amazon SageMaker AI Ground Truth Plus, meninjau label, dan memenuhi prasyarat SageMaker AI Ground Truth Plus.

[Jelajahi panduannya”](#)

- Mulailah dengan Amazon Ground Truth Watch cara memulai pelabelan data Anda dalam hitungan menit melalui konsol SageMaker AI Ground Truth.

[Tonton videonya”](#)

- Amazon SageMaker AI Ground Truth Plus — buat kumpulan data pelatihan tanpa kode atau sumber daya internal

Pelajari tentang Ground Truth Plus, layanan turn-key yang menggunakan tenaga kerja ahli untuk memberikan kumpulan data pelatihan berkualitas tinggi dengan cepat, dan mengurangi biaya hingga 40 persen.

[Baca blog”](#)

SageMaker AI JumpStart

- Memulai pembelajaran mesin dengan SageMaker AI JumpStart

Jelajahi template JumpStart solusi SageMaker AI yang menyiapkan infrastruktur untuk kasus penggunaan umum, dan contoh notebook yang dapat dieksekusi untuk pembelajaran mesin dengan AI. SageMaker

[Jelajahi panduannya”](#)

- Memulai proyek pembelajaran mesin Anda dengan cepat menggunakan Amazon SageMaker AI JumpStart

Pelajari cara mempercepat proyek ML Anda menggunakan model terlatih dan solusi bawaan yang ditawarkan oleh Amazon AI. SageMaker JumpStart Anda kemudian dapat menerapkan model yang dipilih melalui notebook Amazon SageMaker AI Studio.

[Mulai dengan tutorial”](#)

- Dapatkan langsung dengan Amazon SageMaker AI JumpStart dengan lokakarya Immersion Day ini

Pelajari bagaimana kemampuan HTML kode rendah yang ditemukan di Amazon SageMaker AI Data Wrangler, Autopilot, dan Jumpstart, membuatnya lebih mudah untuk bereksperimen lebih cepat dan menghadirkan model yang sangat akurat ke produksi.

[Mulai dengan Workshop”](#)

SageMaker AI Pipelines

- Memulai dengan Amazon SageMaker AI Pipelines

Pelajari cara membuat end-to-end alur kerja yang mengelola dan menerapkan pekerjaan SageMaker AI. SageMaker AI Pipelines hadir dengan integrasi SageMaker AI Python SDK, sehingga Anda dapat membangun setiap langkah pipeline menggunakan antarmuka berbasis Python.

[Jelajahi panduannya”](#)

- Mengotomatiskan alur kerja pembelajaran mesin

Pelajari cara membuat dan mengotomatiskan alur kerja end-to-end machine learning (ML) menggunakan Amazon SageMaker AI Pipelines, Amazon AI Model Registry, dan Amazon SageMaker SageMaker AI Clarify.

[Mulai dengan tutorial”](#)

- Cara membuat alur kerja HTML yang sepenuhnya otomatis dengan Amazon SageMaker AI Pipelines

Pelajari tentang Amazon SageMaker AI Pipelines, CI/CD layanan ML pertama di dunia yang dirancang agar dapat diakses oleh setiap pengembang dan ilmuwan data. SageMaker AI Pipelines membawa CI/CD jaringan pipa ke ML, mengurangi waktu pengkodean yang diperlukan.

[Tonton videonya”](#)

SageMaker AI Studio

- Membangun dan melatih model pembelajaran mesin secara lokal

Pelajari cara membuat dan melatih model ML secara lokal dalam notebook Amazon SageMaker AI Studio Anda.

[Mulai dengan tutorial”](#)

- SageMaker Integrasi AI Studio dengan bengkel EMR

Pelajari cara memanfaatkan pemrosesan terdistribusi dalam skala besar untuk menyiapkan data dan selanjutnya melatih model pembelajaran mesin.

[Mulai dengan Workshop”](#)

Kumpulan Layanan AWS jalur AI/ML ketiga menyediakan dokumentasi mendalam, tutorial langsung, dan sumber daya untuk memulai, dan Amazon Titan. AWS Trainium AWS Inferentia

AWS Trainium

- Menskalakan pelatihan terdistribusi dengan AWS Trainium dan Amazon EKS

Pelajari bagaimana Anda dapat memperoleh manfaat dari ketersediaan umum instans Amazon EC2 Trn1 yang didukung oleh AWS Trainium—akselerator HTML yang dibuat khusus yang dioptimalkan untuk menyediakan platform berkinerja tinggi, hemat biaya, dan skalabel besar-besaran untuk melatih model pembelajaran mendalam di cloud.

[Baca blog”](#)

- Ikhtisar AWS Trainium

Pelajari tentang AWS Trainium, akselerator pembelajaran mesin generasi kedua (ML) yang dirancang khusus AWS untuk pelatihan pembelajaran mendalam model parameter 100B+. Setiap instans Amazon Elastic Compute Cloud (EC2) Trn1 menggunakan hingga 16 AWS Trainium akselerator untuk menghadirkan solusi berkinerja tinggi dan berbiaya rendah untuk pelatihan deep learning (DL) di cloud.

[Jelajahi panduannya](#)

- Instans Trainium yang Direkomendasikan

Jelajahi bagaimana AWS Trainium instance dirancang untuk memberikan kinerja tinggi dan efisiensi biaya untuk beban kerja inferensi model pembelajaran mendalam.

[Jelajahi panduannya](#)

AWS Inferentia

- Ikhtisar AWS Inferentia

Pahami bagaimana akselerator dirancang AWS untuk memberikan kinerja tinggi dengan biaya terendah untuk aplikasi inferensi deep learning (DL) Anda.

[Jelajahi panduannya](#)

- AWS Inferentia 2 dibangun di atas AWS Inferentia 1 dengan memberikan throughput 4x lebih tinggi dan latensi 10x lebih rendah

Pahami untuk apa AWS Inferentia 2 dioptimalkan - dan jelajahi bagaimana ia dirancang dari bawah ke atas untuk memberikan kinerja yang lebih tinggi sekaligus menurunkan biaya LLMs dan inferensi AI generatif.

[Baca blog](#)

- Inferensi pembelajaran mesin menggunakan AWS Inferentia

Pelajari cara membuat kluster Amazon EKS dengan node yang menjalankan instans Amazon EC2 Inf1 dan (opsional) menerapkan aplikasi sampel. Instans Amazon EC2 Inf1 didukung oleh

AWS Inferentia chip, yang dibuat khusus AWS untuk memberikan kinerja tinggi dan inferensi biaya terendah di cloud.

[Jelajahi panduannya](#)

Amazon Titan

- Sekilas tentang Amazon Titan

Jelajahi bagaimana Amazon Titan FMs dilatih sebelumnya pada kumpulan data besar, menjadikannya model yang kuat dan serba guna. Pelajari bagaimana Anda dapat menggunakannya apa adanya - atau secara pribadi - untuk menyesuaikannya dengan data Anda sendiri untuk tugas tertentu tanpa membuat anotasi volume data yang besar.

[Jelajahi panduannya](#)

Jelajahi

- Diagram arsitektur

Diagram arsitektur referensi ini menunjukkan contoh layanan AWS AI dan ML yang digunakan.

[Jelajahi diagram arsitektur](#)

- Whitepaper

Jelajahi whitepaper untuk membantu Anda memulai dan mempelajari praktik terbaik dalam memilih dan menggunakan AI/ML layanan.

[Jelajahi whitepaper](#)

- AWS Solusi

Jelajahi solusi yang diperiksa dan panduan arsitektur untuk kasus penggunaan umum untuk layanan AI dan ML.

[Jelajahi solusi](#)

Sumber daya

Model pondasi

Model pondasi yang didukung meliputi:

- [Antropik Claude](#)
- [Cohere Perintah & Sematkan](#)
- [AI21 Lab Jurassic](#)
- [Meta Llama](#)
- [Mistral AI](#)
- [Difusi Stabil XL](#)
- [Amazon Titan](#)

Menggunakan Amazon Bedrock, Anda dapat bereksperimen dengan berbagai model pondasi dan menyesuaikannya secara pribadi dengan data Anda.

Kasus penggunaan atau layanan khusus industri

- [Amazon Comprehend Medical](#)
- [Amazon Fraud Detector](#)
- [AWS HealthLake](#)
- [Amazon Lookout for Equipment](#)
- [Amazon Lookout for Metrics](#)
- [Amazon Lookout for Vision](#)
- [Amazon Monitron](#)
- [AWS HealthOmics](#)
- [AWS Panorama](#)

Posting blog terkait

- [Kemampuan baru yang signifikan memudahkan penggunaan Amazon Bedrock untuk membangun dan menskalakan aplikasi AI generatif — dan mencapai hasil yang mengesankan](#)
- [AWS Inferensia dan AWS Trainium memberikan biaya terendah untuk menerapkan model Llama 3 di Amazon AI SageMaker JumpStart](#)

- [Merevolusi Kepuasan Pelanggan dengan model hadiah yang disesuaikan untuk bisnis Anda di Amazon AI SageMaker](#)
- [Amazon Personalize meluncurkan resep baru yang mendukung katalog item yang lebih besar dengan latensi lebih rendah](#)

Riwayat dokumen

Tabel berikut menjelaskan perubahan penting pada panduan keputusan ini. Untuk pemberitahuan tentang pembaruan panduan ini, Anda dapat berlangganan umpan RSS.

Perubahan	Deskripsi	Tanggal
Pembaruan kecil	Konten yang diperbarui untuk Amazon Q dan tumpukan AI dan ML terbaru Amazon.	3 Mei 2024
Rilis awal	Rilis awal panduan keputusan.	Juli 24, 2023

Terjemahan disediakan oleh mesin penerjemah. Jika konten terjemahan yang diberikan bertentangan dengan versi bahasa Inggris aslinya, utamakan versi bahasa Inggris.