BUKU PANDUAN
PROGRAM: ANALITIK
NOMOR DOKUMEN: T147

NUCLEUS RESEARCH

BUKU PANDUAN
DEEP LEARNING DI AWS

ANALIS
Daniel Elman

Nucleus Research, Inc.
100 State Street, Boston, MA, 02109
+1 (617) 720-2000
www.nucleusresearch.com
©2019 Nucleus Research, Inc.

HASIL AKHIR

Deep learning masih menjadi topik terhangat dalam area kecerdasan buatan (AI) dan bergerak dengan cepat dari akademi serta teori menuju workload yang dioperasikan dan memiliki nilai tambah. Kemajuan dalam infrastruktur data dan kemampuan komputasi, serta kelas baru dari neural network telah membantu membuat deep learning menjadi layak untuk dimanfaatkan oleh bisnis modern, yang hanyalah sebuah mimpi kecil satu tahun lalu. Bekerja sama dengan vendor cloud seperti Amazon Web Services (AWS) dan memanfaatkan layanan machine learning yang di-host cloud seperti Amazon SageMaker telah menjadi kunci bagi perusahaan yang ingin mempercepat proyek deep learning dari konsep hingga operasional. Di masa lalu, ketika infrastruktur belum secanggih saat ini dan layanan machine learning belum matang atau belum tersedia, organisasi yang tidak memiliki anggaran dan keahlian untuk mengembangkan sistem AI dari dasar secara internal mulai tertinggal ketika perusahaan lebih besar dengan lebih banyak anggaran serta pengembang mulai meneliti juga deploy kemampuan AI tersebut.

Untuk memahami kondisi pengadopsian dan penggunaan deep learning saat ini, serta bagaimana perubahannya dibanding laporan tahun lalu (Nucleus Research *s180 – Guidebook: TensorFlow on AWS* –November 2018), Nucleus mengadakan wawancara dan menganalisis pengalaman dari 316 proyek yang unik. Untuk kedua kalinya dalam dua tahun, jumlah proyek deep learning dalam tahap operasional bertambah lebih dari dua kali lipat. Kami menemukan bahwa 96 persen deep learning yang dijalankan di lingkungan cloud, dengan TensorFlow sebagai framework paling populer, digunakan dalam 74 persen proyek deep learning. PyTorch juga digunakan dalam 43 persen proyek (harap diingat bahwa kebanyakan proyek memanfaatkan beberapa framework; MXNet, Keras, dan Caffe2 juga muncul, pada dasarnya selalu berhubungan dengan TensorFlow, PyTorch, atau keduanya), sebuah peningkatan pengadopsian yang signifikan dibanding dengan tahun lalu. Dari 316 proyek, hanya 9 persen yang dibangun dengan satu framework. Hal yang paling menonjol, dari proyek deep learning yang di-host cloud, adalah 89 persen dijalankan di AWS; pendorong utama dari hal tersebut adalah keberagaman pilihan framework di AWS serta investasi berkelanjutannya pada layanan yang baru serta yang sudah ada.

Kami juga menemukan bahwa 85 persen beban kerja TensorFlow berbasis cloud dijalankan di AWS, dan 83 persen proyek PyTorch berbasis cloud ada di AWS. Tahun lalu, sekitar sepertiga pelanggan yang diwawancara mempertimbangkan atau menggunakan SageMaker, layanan yang dikelola Amazon untuk build, train, deploy, dan merancang model deep learning; berdasarkan pengguna yang diwawancara tahun ini, 63 persen pelanggan Amazon telah mulai menggunakan SageMaker.

SITUASI

Machine learning meliputi teknologi di mana komputer menganalisis data untuk "belajar" dari pengalaman tanpa keterlibatan manusia. Deep learning adalah subset machine learning; dalam machine learning, komputer diberi data yang memiliki kumpulan fitur yang telah ditentukan untuk dianalisis. Sedangkan pada deep learning, komputer diberi data tidak terstruktur seperti teks, audio, atau video, dan komputer menentukan sendiri manakah fitur yang relevan dengan analisis. Singkatnya, komputer diberikan sepasang input sampel serta output yang sesuai dan dapat bekerja mundur untuk menemukan operasi apa yang diperlukan untuk mengubah input menjadi output.

Model deep learning modern memerlukan jumlah komputasi dan penyimpanan yang sangat besar, mempersulit banyak organisasi untuk build sendiri sistem ini. Jadi, seperti yang kami temukan selama riset ini, perusahaan sangat ingin memanfaatkan cloud untuk proyek deep learning. Pendekatan ini memungkinkan bisnis untuk membeli jumlah penyimpanan data dan kemampuan komputasi yang mereka perlukan untuk proyek tanpa perlu membeli, mengonfigurasi, dan memelihara infrastruktur secara internal, menghemat biaya secara signifikan dari waktu ke waktu.

Dengan basis pengguna yang beragam dan berkembang cepat berkat tren serta potensi AI, lanskap teknologi berubah dengan cepat. Alat dan metodologi baru mulai tersedia secara terus-menerus. Oleh karena itu, mendukung pengguna untuk mengembangkan pada platform dengan fleksibilitas maksimal merupakan kunci bagi kesuksesan jangka panjang di pasar machine learning berbasis cloud. Singkatnya, platform cloud harus mendukung beragam alat dan framework pengembangan yang digunakan saat ini juga di masa mendatang, dengan keamanan serta ketersediaan yang diperlukan dalam mematuhi peraturan penanganan data dan privasi.

Kami telah melakukan studi ini selama tiga tahun berturut-turut, dan selama tiga tahun ini kami telah melihat perubahan transformatif dalam kemampuan model, kekuatan komputasi, dan alat pengembang dalam mendukung hasil baru yang menarik. Di tahun pertama, kami sulit menemukan organisasi yang telah bergerak melampaui tahap pengembangan awal dan proyek pembuktian konsep dengan deep learning. Pada 2018, kami melihat kemajuan yang kuat dengan 14 persen proyek diklasifikasikan sebagai dalam tahap operasional, yang menangani data langsung. Lompatan baru pada efek tersebut terjadi di tahun ini, di mana organisasi yang meliputi perusahaan rintisan dengan 20 karyawan hingga perusahaan global Fortune 100 deploy deep learning ke tahap operasional, dengan 38 persen proyek dalam tahap operasional. Aspek lain yang didiskusikan dalam wawancara meliputi:

- Sasaran dan motivasi di balik proyek deep learning
- Strategi penerapan dan manfaat yang terkait
- Framework pengembangan, metode, dan alat lain yang digunakan
- Kekuatan dan kelemahan dari beragamnya model serta framework

- Jumlah orang yang terlibat dan perannya masing-masing dalam tim proyek

Secara total, Nucleus melakukan wawancara mendalam dengan 32 ahli deep learning, yang kebanyakan bertanggung jawab atas beberapa proyek secara bersamaan, mewakili 316 proyek yang unik.

DEEP LEARNING DI CLOUD

Beban kerja deep learning skala produksi melibatkan pemrosesan ribuan atau jutaan data contoh untuk train model. Ini sangat mahal dari segi komputasi, khususnya untuk data input yang rumit seperti gambar atau video, dan kebanyakan organisasi tidak sanggup build serta memelihara sistem komputasi berkinerja tinggi dengan CPU atau GPU yang diparalelkan untuk melakukan kalkulasi. Hasilnya, organisasi memilih cloud untuk mengakses sumber daya dan infrastruktur yang mereka butuhkan. Tahun ini, kami menemukan bahwa 96 persen proyek deep learning dijalankan di lingkungan cloud—ini mencerminkan temuan tahun lalu, namun dengan 177 proyek di 2018 berkembang menjadi 316 di 2019 tetap menunjukkan momentum pelanggan yang kuat terhadap cloud untuk deep learning. Dari beban kerja yang ada di tahap operasional pada data langsung, 98 persen berada di cloud. Untuk organisasi yang belum sepenuhnya berada di cloud, strategi penerapan yang umum melibatkan pengembangan model skala kecil pada perangkat keras di lokasi sebelum bermigrasi ke cloud untuk produksi.

MANFAT CLOUD

Dengan 96 persen proyek deep learning dijalankan di cloud, jelas sekali bahwa pelanggan memahami nilai pendekatan ini. Kami meminta responden untuk mengidentifikasi manfaat utama dari menjalankan deep learning di cloud. Respons mereka berkisar di antara tiga tema utama:

- Hemat biaya karena tidak perlu membayar biaya perangkat keras, personel, dan energi. Ini adalah repons yang paling umum, dikutip dari setiap peserta wawancara. Deep learning memerlukan komputasi dalam jumlah besar; build dan memelihara sistem perangkat keras yang dapat melakukan deep learning yang sesuai memerlukan profesional TI khusus; menjalankan perangkat keras secara fisik untuk train model deep learning menghabiskan ribuan jam waktu CPU dan GPU—selain biaya listrik yang juga seringkali menjadi hambatan. Dengan cloud, pengguna hanya membayar sumber daya yang mereka gunakan tanpa biaya terkait.
- Kemampuan untuk berkolaborasi dan bekerja dalam tim yang terdistribusi. Model yang diterapkan di cloud dapat diakses oleh seluruh pengguna yang mendapat izin, terlepas dari lokasi fisik mereka. Hal ini mempercepat pengembangan model, khususnya di seluruh tim jarak jauh yang menjadi semakin umum.

- Kemampuan untuk memanfaatkan kapabilitas dan alat platform tambahan. Keamanan dan ketersediaan merupakan aspek yang paling umum disebutkan. Sistem cloud yang dikonfigurasi dengan tepat mendapat manfaat dari investasi keamanan pelanggan dan penyedia cloud. Selain itu, kemampuan untuk menjalankan model di pusat data lokal agar tetap mematuhi undang-undang perlindungan data seperti GDPR merupakan hal yang penting. Selain penyimpanan dan komputasi, vendor cloud menawarkan kemampuan alat serta platform untuk meningkatkan pengalaman pengembang. Alat seperti Amazon SageMaker memberikan nilai yang hebat kepada pelanggan cloud dengan alur kerja machine learning menyeluruh yang dikelola sepenuhnya - dari membersihkan data hingga train, build, dan deploy model.

DEEP LEARNING ADALAH HAL NYATA

Sebuah langkah maju pada kondisi machine learning telah ditunjukkan di tahun lalu, dengan 14 persen proyek dalam tahap operasional. Kami melihat lompatan yang serupa di 2019 dengan 38 persen dari 316 proyek deep learning diklasifikasikan sebagai dalam tahap operasional. 89 persen proyek deep learning yang dalam tahap operasional dijalankan di AWS. 76 persen proyek yang dalam tahap operasional memanfaatkan TensorFlow dan 28 persen proyek menggunakan PyTorch. Keras dan Apache MXNet juga terlihat pada pengaturan tahap operasional karena kebanyakan proyek memiliki komponen yang dibangun dengan beberapa framework. Hanya 9 persen proyek yang dibangun menggunakan satu framework saja. Ketika perusahaan menyadari bahwa deep learning dan kemampuan AI lainnya telah mencapai level kematangan yang memungkinkan mereka memberikan nilai bisnis yang logis, mereka mulai deploy “keuntungan yang diperoleh dengan mudah”, atau kasus penggunaan yang lebih sederhana yang didemonstrasikan dengan baik dan cepat untuk diimplementasikan. Contoh yang umum meliputi antarmuka suara untuk situs web dan aplikasi serta mesin rekomendasi untuk situs belanja online.

Perusahaan juga masih menjelajahi aplikasi deep learning yang lebih rumit, namun banyak yang masih dalam tahap pengujian. Hasil studi tahun ini menunjukkan bahwa perusahaan memperluas seluruh investasi deep learning mereka sejak tahun lalu, melanjutkan kemajuan pada proyek aspirasional yang lebih rumit selama beberapa tahun, seraya menambahkan aplikasi yang dapat diterapkan dengan cepat dan memiliki nilai tambah seperti rekomendasi, analisis sentimen di chatbot, dan antarmuka suara, agar dapat mengikuti tren pasar juga menunjukkan keberlanjutan nilai serta kelayakan deep learning pada penggunaannya di dunia nyata.

DEEP LEARNING DI AWS

Nucleus menemukan bahwa alasan utama dalam memilih AWS—kemampuan platform yang beragam, hubungan dengan Amazon, dan investasi berkelanjutan AWS dalam layanan deep learning—tetap tidak berubah sejak tahun lalu.

ANEKA RAGAM KEMAMPUAN AMAZON

Amazon mendukung proses deep learning secara menyeluruh, sehingga tidak perlu menggabungkan komponen terbaik dari beragam vendor. Pelanggan dapat menyimpan data, build dan deploy model, serta membuat aplikasi yang memanfaatkan output model seluruhnya di platform. Pelanggan dapat bebas memilih perangkat keras khusus yang dioptimalkan untuk beban kerja mereka; contohnya, mereka dapat mengakses instans GPU canggih yang dioptimalkan untuk deep learning, seperti Amazon EC2 P3 dan G4, sesuai permintaan. Amazon memiliki pusat data wilayah di seluruh dunia, sehingga pelanggan dapat melokalkan data dan pengoperasian mereka sesuai kebutuhan serta mematuhi peraturan pembagian data wilayah. Keamanan level platform menambahkan lapisan perlindungan pada data dan aplikasi. Alat dan layanan lain yang terintegrasi di platform seperti layanan container meminjamkan dirinya ke proyek berskala besar. Berikut ini adalah komentar dari para pengguna:

- *"Penyedia cloud lain tidak dapat menandingi kematangan dan kelengkapan platform AWS. Melalui kemitraan, kami mendapat lebih dari sekadar teknologi; mereka memiliki keahlian dan layanan tambahan seperti S3 dan CloudFormation untuk membantu kami mengubah proyek teknologi dari konsep menjadi produksi."*
- *"Kami terus mengevaluasi vendor cloud lain, namun sejauh ini belum ada yang dapat menandingi AWS. Belum ada vendor yang memberikan nilai yang sebanding dalam layanan cloud, dan ini belum termasuk keamanan juga ketersediaan platform terbaik di kelasnya."*

Profil pengguna - Perusahaan bioteknologi

Sebuah perusahaan bioteknologi menggunakan deep learning untuk mengembangkan pengklasifikasi tingkat produksi guna memprediksi kanker dari data genomik. Karena masih dalam tahap riset, perusahaan build model untuk menguji beragam hipotesis secara paralel, sehingga fleksibilitas untuk deploy dan menguji ide baru dengan cepat merupakan hal yang sangat penting.

Terdapat empat pendekatan berbeda yang diuji secara paralel; dua dibangun di TensorFlow, satu di PyTorch, dan satu lagi dikembangkan dengan kode yang disesuaikan seluruhnya. Framework dipilih secara strategis—TensorFlow dipilih untuk dua proyek estimator dan lost function yang tertanam sebelumnya; PyTorch dipilih karena pendekatannya terinspirasi oleh proyek di akademia yang berdasarkan pada PyTorch.

Perusahaan memiliki sejumlah alasan dalam memilih AWS untuk proyek deep learning mereka. Amazon menawarkan beragam opsi komputasi yang sangat sesuai untuk aneka

kasus penggunaan. Deep learning mengenai data genomik memerlukan pra-pemrosesan yang ekstensif, sehingga perusahaan memerlukan mesin yang dioptimalkan agar memori dapat melakukan pra-pemrosesan data train, dan mesin yang diakselerasi GPU untuk train model. Bekerja dengan data genomik train yang dalam ukuran petabyte, kemampuan mitra cloud untuk mendukung proyek pada skala tersebut sangatlah krusial. Karena proyek menggunakan framework yang berbeda, penting agar penyedia cloud mendukung beberapa framework sebagai penghuni utama pada infrastrukturnya tanpa memerlukan pengaturan yang ekstensif. Amazon sangat cocok karena kinerja yang ditawarkannya, bahkan pada skala petabyte, dan fleksibilitas mesin yang didukung serta uasahanya untuk mengoptimalkan seluruh framework agar dapat berfungsi dengan baik.

HUBUNGAN DENGAN AMAZON

Pelanggan yang diwawancara juga membahas hubungan mereka dengan AWS sebagai faktor pendorong di balik keputusan bisnis. Banyak pelanggan yang telah menggunakan AWS di area bisnis lain dan menumbuhkan investasi mereka untuk menyertakan deep learning. Deep learning memerlukan banyak data, dan organisasi telah memercayakan data mereka ke AWS. Amazon juga memberikan teknologi dan praktik terbaik kepada pelanggan guna menyelesaikan proyek mereka tanpa memaksakan solusi tertentu kepada mereka. Pelanggan dengan cepat menyadari fleksibilitas platform—mereka dapat menggunakan framework dan pustaka yang paling nyaman bagi mereka tanpa perlu merasa khawatir bahwa hal tersebut tidak sesuai dengan teknologi AWS. Satu pelanggan menyebutkan bahwa ketertarikan Amazon terhadap studi seperti ini menunjukkan komitmen Amazon dalam memaksimalkan nilai pelanggan. Pengguna lainnya berkata:

- *"Kami build seluruh bisnis kami di AWS, jadi kami tidak akan mudah untuk berpindah ke penyedia lain. [Arsitektur] AWS kami telah tumbuh bersama kami, dan dukungannya dalam setiap langkah kami tidak perlu diragukan. Kami sungguh merasa bahwa mereka berinvestasi dalam kesuksesan kami, dan ini merupakan kunci dari setiap hubungan jangka panjang."*
- *"Dengan Amazon, kami memiliki akses ke kasus penggunaan pelanggan dan keahlian industri bagi proyek machine learning kami. Tim dukungan sangat berjasa dalam upaya kami—mereka dapat menjawab seluruh pertanyaan kami terkait framework, pilihan model, dan persyaratan infrastruktur."*

Profil pengguna - Perusahaan Media Digital dan Game Seluler

Sebuah perusahaan media digital yang merancang dan menjual game seluler menggunakan machine learning untuk menyeimbangkan kesulitan game dengan pendapatan terproyeksi yang dihasilkan. Ini dilakukan dengan menggunakan pembelajaran dukungan mendalam untuk train bot memainkan setiap game. Lalu kinerja bot dipantau guna memperkirakan kesulitan di tiap level yang kemudian digunakan untuk memprediksi berapa banyak pengguna yang akan berhenti bermain di tiap level. Tentu

saja, ketika pengguna berhenti bermain, mereka berhenti menghasilkan pendapatan iklan untuk vendor game, sehingga perusahaan menggunakan proses ini untuk menyeimbangkan kesulitan game dengan pendapatan terproyeksi.

AWS telah menjadi penyedia cloud bagi perusahaan sejak 2008, jadi tentu saja mereka menggunakan deep learning AWS. Kedua proyek dibangun pada PyTorch—PyTorch merupakan standar industri bagi pembelajaran dukungan mendalam, dan banyak riset terbaru diterbitkan dengan menyertakan implementasi PyTorch.

Perusahaan menggunakan SageMaker untuk deploy model yang bertugas memprediksi pengurangan pemain ke dalam produksi. Ini dipilih karena memberikan "pandangan menyeluruh" kepada kepemimpinan proyek mengenai model dan lokasi terpusat untuk melihat serta mengendalikan seluruh model dalam penerapan. Dengan teknologi seperti SageMaker dikombinasikan bersama sumber daya komputasi yang ditawarkan Amazon, ahli yang diwawancara berujar, "memerlukan kasus bisnis yang menarik untuk membujuk kami agar meninggalkan AWS dan sejauh ini belum ada penyedia cloud yang mampu menawarkan kami alat khusus machine learning serta penyimpanan juga komputasi untuk [nilai yang lebih baik]."

INVESTASI AWS DI DEEP LEARNING

Pelanggan tahu bahwa Amazon mengembangkan dan menggunakan teknologi deep learningnya sendiri. Para ahli deep learning menyebutkan bahwa peningkatan berkelanjutan pada dokumentasi, dukungan framework, dan layanan cloud seperti Amazon SageMaker sebagai faktor utama dalam memilih AWS dibanding penyedia cloud lainnya. SageMaker mulai tersedia pada 2017; ini adalah layanan cloud yang dikelola sepenuhnya dan mencakup seluruh alur kerja machine learning - dari build, train, hingga deploy model machine learning.

Pengadopsian SageMaker tumbuh dengan cepat karena pengembang menyadari bahwa ini dapat mengurangi kerumitan dan mempercepat penerapan model. Tahun lalu, sekitar sepertiga responden menggunakan atau menjelajahi penggunaan SageMaker untuk mengautomasi aspek pada proyek deep learning mereka; tahun ini angka tersebut meningkat hampir dua kali lipat dengan 63 persen pelanggan menggunakan atau mempertimbangkan SageMaker. Selama studi ini, kami menemukan pelanggan yang memigrasi penerapan TensorFlow yang mereka kembangkan sendiri ke layanan terkelola di AWS melalui SageMaker, serta pelanggan yang build seluruh sistem mereka dari dasar dengan SageMaker. Berikut ini adalah komentar dari para pengguna:

- *"Kami tidak perlu menggunakan perangkat lunak khusus untuk menjalankan proyek deep learning berskala besar. Tanpa AWS dan khususnya SageMaker, kami pasti harus membeli perangkat keras, train model secara lokal, lalu menyimpan dan me-host model di server internal agar dapat diakses saat saya menginginkannya untuk melakukan forecasting. Untuk memulainya saja memerlukan waktu berminggu-minggu dan harus mengeluarkan banyak biaya tambahan untuk perangkat keras,*

listrik, dan personel, belum lagi seluruh waktu yang hilang untuk build kemampuan secara internal yang sebenarnya sudah ada di SageMaker."

- *"Kami berfokus untuk deploy machine learning konversasional bagi para klien kami dan SageMaker menyederhanakan pembangunan modelnya. Rata-rata, kami memiliki lima proyek yang sedang berjalan untuk satu klien pada satu pekan tertentu. Sejak menggunakan SageMaker tahun lalu, kami menemukan bahwa kami menghemat sekitar dua jam per proyek dengan mengautomasi train model terdistribusi."*

Profil pengguna – Perusahaan Bisnis Perangkat Lunak

Sebuah perusahaan bisnis perangkat lunak global yang memproduksi aplikasi khusus untuk penjualan dan tim layanan mengadopsi Amazon SageMaker guna mengelola penerapan TensorFlow-nya. Upaya deep learningnya umumnya seputar analisis dan klasifikasi interaksi pelanggan guna memahami secara efektif bagaimana beragam jenis jangkauan memengaruhi kecenderungan pelanggan untuk kembali atau membeli lagi.

Sebagai perusahaan teknologi besar, perusahaan ini berhasil mengungguli tren saat ini dengan upaya deep learningnya dibandingkan pasar yang lebih besar. Sebelum SageMaker, perusahaan build secara khusus sebagian besar infrastruktur deep learning berbasis TensorFlow. Di awal tahun, perusahaan memutuskan untuk memigrasi penerapan TensorFlow yang dikelola sendiri ke SageMaker yang dapat dikelola sebagai sebuah layanan. Sementara upaya terus berlangsung dan belum selesai sepenuhnya, organisasi telah mampu menetapkan kembali 3 FTE yang secara khusus bertanggung jawab untuk mengelola ekosistem TensorFlow. Selain itu, kecepatan model train meningkat secara drastis karena SageMaker mendistribusikan muatan komputasi secara otomatis ke beberapa CPU dan GPU secara paralel. Perusahaan melaporkan bahwa penerapan model baru dengan SageMaker hanya memerlukan kurang dari 50 persen waktu yang diperlukan untuk melakukannya di lingkungan yang dikelola sendiri.

Profil pengguna – Perusahaan Pengembangan Aplikasi

Sebuah perusahaan pengembangan aplikasi yang khusus menciptakan game terintegrasi dengan suara yang dapat dimainkan di Amazon Alexa, si speaker cerdas, build seluruh proyek deep learning di Amazon SageMaker yang menganjurkan game untuk menjaga keterlibatan pengguna agar tetap tinggi. Sistem menggunakan data dari game yang dimainkan sebelumnya pada sistem untuk secara prediktif merekomendasikan game lain yang serupa kepada pengguna.

Perusahaan build bisnisnya di platform AWS, sehingga SageMaker dipilih untuk integrasi asli dengan arsitektur yang ada di AWS, khususnya Amazon S3 dan AWS Lambda, untuk mengakses data tersimpan dan komputasi tanpa server. Selain itu, sejak perusahaan menggunakan AWS, user permission dan prosedur devops telah diformalkan, sehingga

memungkinkan mereka untuk menghindari duplikasi upaya tersebut. Train model yang diparalelkan dan UI level kontrol membuat train serta evaluasi menjadi jauh lebih cepat daripada yang dilakukan secara manual. Pelanggan memperkirakan bahwa penggunaan SageMaker membuat train model tiga kali lebih cepat dan penerapan model empat kali lebih cepat dibanding sistem yang dikelola secara manual.

KESIMPULAN

Tahun lalu, kami melihat proyek deep learning bergerak keluar dari ruang kelas secara besar-besaran menuju ruang dewan direksi karena kemajuan dalam edukasi machine learning, teknologi komputasi, infrastruktur cloud, dan kumpulan data yang sesuai membantu deep learning tersedia secara komersil. Tahun ini, tren terus berlanjut dengan semakin banyaknya perusahaan yang ingin deploy deep learning yang sesuai guna menyelesaikan masalah di dunia nyata. Untuk kedua kalinya dalam dua tahun, jumlah proyek deep learning dalam tahap operasional bertambah lebih dari dua kali lipat, umumnya karena faktor utama yang sama seperti tahun lalu.

- Meningkatnya ketersediaan infrastruktur dan layanan cloud dari vendor seperti AWS untuk mendukung proses komputasi intensif dengan data yang besar seperti deep learning.
- Peningkatan menuju deep learning yang canggih dengan teknik, arsitektur jaringan, dan kumpulan data yang ditingkatkan membuat jaringan neural menjadi lebih akurat dan andal.
- Investasi berkelanjutan di komunitas untuk membagikan pengalaman dan mendukung peneliti deep learning lain melalui forum online dan dokumentasi, pustaka serta framework sumber terbuka, juga penawaran cloud seperti model internal serta perangkat keras khusus yang dioptimalkan untuk machine learning.

Karena berkurangnya biaya awal untuk menjelajahi deep learning, kami melihat semakin banyak bisnis yang ingin bergabung. Daripada mencari cara untuk menemukan kembali penggeraknya, strategi yang paling efektif hingga saat ini adalah bermitra dengan vendor cloud yang memiliki infrastruktur, keahlian, dan layanan tambahan untuk mengubah deep learning, dari konsep hingga selesai. Menurut analisis kami, reputasi Amazon sebagai penyedia teknologi cloud bisnis paling matang dan canggih beserta investasinya yang berfokus pada bidang dalam layanan machine learning serta fleksibilitas platform untuk mendukung pilihan arsitektur jaringan, framework pengembangan, atau sumber data pelanggan menjadikannya platform cloud pilihan bagi profesional deep learning.