



IDC EXECUTIVE SUMMARY

Accelerate Machine Learning Development to Build Intelligent Applications Faster

Sponsored by Amazon Web Services

Author: David Schubmehl



Sponsored by:
Amazon Web Services

Author: David Schubmehl

October 2020

Highlights

By 2021, spending on AI applications will exceed
\$63 BILLION

By 2023, growing to over
\$96 BILLION

And by 2025, at least
90%
of new enterprise application releases will include embedded AI-based functionality, recommendations, or advice

Accelerate Machine Learning Development to Build Intelligent Applications Faster

INTRODUCTION

Organizations are experiencing situations and business conditions unlike anything they have seen in the past 50 years. The COVID-19 pandemic and related business conditions are leading companies to implement digital transformation and process reorganization at a rate that would have been unthinkable six months ago. To implement digital transformation and process reorganization successfully, companies are working to improve and provide better value to their customers, shareholders, and workers. Specifically, they are looking for new ways to increase sales, reduce costs, streamline business processes, and understand their customers better. Artificial intelligence (AI), machine learning (ML), and deep learning (DL) are key technologies that help organizations actualize digital transformation. Because of this scenario, IDC is seeing that many organizations are accelerating deployments and changing the focus of their AI projects from proof of concept (POC) to putting applications in production that deliver return on investment (ROI).

Organizations are seeing AI/ML/DL technologies as effective mechanisms to get closer to their customers, improve business processes, and reduce costs. Organizations, data scientists, and developers are feeling the pressure from their management teams, boards, and even customers to grab on to this new wave of computing. However, as with many emerging technologies, critical success factors around ML/DL implementations are related to the people, process, and technologies involved. Emerging technical solutions have traditionally required sharp and motivated developers who like to live on the cutting edge of technology. Creating new intelligent applications using AI takes significant time and energy, as well as experienced developers to really build and deploy meaningful ML/DL solutions.

This IDC Executive Summary looks at some of the trends around this nascent technology and how vendors such as Amazon Web Services (AWS) have developed services and tools for ML/DL that allow data scientists and developers to deploy ML/DL models more quickly and easily.

At least 85% of organizations with over 500 employees surveyed had adopted some type of AI-enabled software.

Situation Overview

The market for ML and DL-based AI applications has grown rapidly and continues to surge. IDC estimates that spending on AI applications will exceed \$63 billion by 2021, growing to over \$96 billion by 2023, and by 2025, at least 90% of new enterprise application releases will include embedded AI-based functionality, recommendations, or advice.

IDC research in 2019 showed that at least 85% of organizations with over 500 employees surveyed had adopted some type of AI-enabled software and the vast majority of those were customizing or developing their own AI/ML models and applications.

According to IDC's *2019 AI Frameworks, Tools, Methods Survey*, 37.5% of organizations are developing their models off premises in a cloud environment. However, public cloud will outweigh on-premises development and deployment by as early as 2023, according to *Worldwide Artificial Intelligence Software Platforms Forecast, 2020–2024* (IDC #US45724520, June 2020). In addition, IDC is seeing more cloud-based projects moving from proof of concept to production than ever before.

This is consistent with the general trends we see with application deployments for other workload types. Developing cloud-based intelligent applications that use DL is increasingly popular across both the enterprise software market and the application development market. IT organizations are feeling pressure from their management teams, boards, and even customers to find advantages and efficiencies with this new wave of computing and the use of increasingly popular public cloud development and deployment.

Even the COVID-19 pandemic has not slowed budgeted IT plans for AI spending. IDC's *COVID-19 Impact on IT Spending Survey* (conducted July 7–20, 2020) indicated that **most organizations are either increasing their AI, ML, and DL spending or leaving it unchanged.** Only 16% of respondents on a worldwide basis planned to reduce AI and DL spending.

To ease development, training, and deployment of deep learning models, AWS has developed and extended Amazon SageMaker into a complete ML and DL development environment.

Given all of this, organizations should be emphasizing the need to invest in and create AI, ML, and DL solutions. However, there are many factors involved that inhibit the development of deep learning models and solutions:

- » Choosing the right deep learning framework for the job at hand
- » Choosing the right deep learning algorithm
- » Adjusting and tuning the deep learning algorithm and data for the most accurate predictions
- » Identifying, locating, and curating training data for deep learning models
- » Having the right amount of compute resources for both model training and generating predictions in production (inferences)
- » Integrating deep learning models into existing enterprise applications
- » Operationalizing models to perform at scale in production

The question for enterprises is how best to develop these ML and DL models while minimizing the amount of effort and time needed to collect and organize data and then develop accurate predictive and prescriptive models.

Considering Amazon SageMaker: A Fully Managed Service for Machine Learning

To ease development, training, and deployment of deep learning models, AWS has developed and extended Amazon SageMaker into a complete ML and DL development environment. SageMaker is a fully managed service that enables developers and data scientists to build, train, and deploy machine learning and deep learning models at any scale quickly and easily. SageMaker removes the complexity that holds back developer success. SageMaker includes modules that can be used together or independently to build, train, and deploy ML and DL models, such as:

- » **Amazon SageMaker Studio:** This is the first fully integrated development environment (IDE) for machine learning where developers and data scientists can perform all development steps. Teams and individuals can quickly upload data, create and share new notebooks, train and tune models, move back and forth between steps to adjust experiments, debug and compare results, and deploy and monitor models (all in a single visual interface), making them much more productive.

Amazon SageMaker requires no setup and provides hosted Jupyter notebooks so that developers can start processing training data sets and developing ML and DL models immediately.

Amazon SageMaker provides the tools and capabilities for organizations to solve complex challenges ranging from pricing optimization and offer management to predictive maintenance and even intelligent process automation.

- » **Amazon SageMaker Autopilot:** Automatically build, train, and tune models with full visibility and control using Amazon SageMaker Autopilot. It incorporates the industry's first automated machine learning capability that gives developers complete control and visibility into how their models were created and what logic was used in creating these models.
- » **Amazon SageMaker Ground Truth:** This fully managed data labeling service makes it easy to build highly accurate training data sets for machine learning. It provides tools and services for labeling data including automatic data labeling, which uses a machine learning model to label first-party data.
- » **Amazon Augmented AI:** This service makes it easy to build the workflows required for human review of ML predictions. It brings human review to all developers, removing the undifferentiated heavy lifting associated with building human review systems or managing large numbers of human reviewers.

SageMaker requires no setup and provides hosted Jupyter notebooks so that developers can start processing training data sets and developing ML and DL models immediately. All it takes is a few clicks in the SageMaker console to create a fully managed notebook workspace. The service takes care of establishing secure network connections to your organization's VPC and launching an Amazon EC2 instance, preloaded with useful libraries for the most popular ML and DL frameworks such as TensorFlow and PyTorch. Developers can build or import their own notebook or just bring data to one of many prebuilt notebooks designed for common use cases, such as risk modeling, churn prediction, and OCR. To prepare the data from Amazon S3, Amazon Redshift, Amazon DynamoDB, and Amazon RDS for model training, developers can use AWS Glue, Apache Spark, or Amazon EMR for data preprocessing, and Amazon EFS as optional storage for your workspace.

SageMaker can be used to solve a wide range of challenges difficult to work out with traditional coding methods ranging from traditional algorithm development to using other forms of ML that may not be as flexible or dynamic as DL. SageMaker allows data scientists and developers to quickly set up, test, and iterate DL/ML solutions at a speed that is not possible with standard ML/DL frameworks and languages and then deploy those solutions into production quickly and easily.

SageMaker provides the tools and capabilities for organizations to solve complex challenges ranging from pricing optimization and offer management to predictive maintenance and even intelligent process automation. Organizations such as GE Healthcare are using SageMaker to improve patient outcomes via the development of DL algorithms that provide predictive and prescriptive diagnoses to their clients and customers. ADP uses Amazon SageMaker to quickly identify workforce patterns

With Amazon SageMaker, developing and deploying a deep learning model can be as straightforward as choosing a notebook template, selecting an algorithm, and then training, testing, and deploying the model using the management service.

and predict outcomes before they happen — for example, employee turnover or the impact of an increase in compensation. By leveraging AWS as its primary platform for artificial intelligence and machine learning, ADP has reduced time to deploy machine learning models from two weeks to just one day.

With SageMaker, developing and deploying a deep learning model can be as straightforward as choosing a notebook template, selecting an algorithm, and then training, testing, and deploying the model using the management service. Amazon also offers a service called Amazon SageMaker Neo that enables developers to train machine learning models developed using popular frameworks once and then run them anywhere in the cloud or at the edge. This service optimizes models to run up to twice as fast as models that are not optimized with Amazon SageMaker Neo, with less than a tenth of the memory footprint, with no loss in accuracy, allowing them to be deployed into almost any environment. The bottom line is that Amazon SageMaker provides an end-to-end machine learning environment that can significantly accelerate and simplify the process of building, training, and deploying models into production applications.

Challenges/Opportunities

COVID-19 is leading every organization to examine and understand its business processes and methods of doing business. The reasons for this are many, but the bottom line is that organizations that are not open to changing the way they do business may not survive this incredibly complex business cycle. Organizations need to understand where AI, machine learning, and deep learning technology will deliver the best business benefits. They also need to understand what skill sets are needed to build and deploy intelligent applications. Finally, organizations need to reassess what the right tools, infrastructure, and environments are to put these intelligent applications to use.

Organizations need guidance about what types of tools and technologies can help them develop intelligent applications.

Organizations need guidance about what types of tools and technologies can help them develop intelligent applications. They also need to understand when, why, and how these applications will be most effective in their organizations. In addition, organizations need to measure the effectiveness of these applications to determine return on investment for future projects that will include machine learning and deep learning.

Offerings such as Amazon SageMaker and the AWS Deep Learning AMLs provide the ways and means for developers to become more productive and deploy deep learning models.

Conclusion

Effective intelligent applications require both advanced technology and solid design judgment. Organizations should make sure the intelligent solution that they are building will be able to help achieve the desired business outcome and/or address the issues that the solution is planned to overcome. Organizations need to engage in-house subject matter experts, the right stakeholders, and consulting partners to help develop the right use cases to align with the desired business outcome. They should include past project experiences in the organization's design thinking approach and, if available, include predefined use cases that have been developed for peers within the organization's industry to help develop the optimal use cases for the desired outcome. This process should involve continuous innovation and prototyping until the right use cases have been developed.

There are a wide variety of tools and libraries available, and it is not always clear which services or libraries are best for the tasks and use cases that developers need to tackle to successfully develop intelligent applications. Offerings such as Amazon SageMaker and the AWS Deep Learning AMLs provide the ways and means for developers to become more productive and deploy deep learning models as well as support services such as data curation, integration, and management to solve a wide range of challenges that are difficult to solve with traditional coding methods and address the organization's business needs. Organizations should be evaluating tools and services like these as they begin to develop and deploy intelligent applications using machine learning and deep learning models.

MESSAGE FROM AMAZON WEB SERVICES

[Read the full white paper](#)

Unlock the Full Potential of Machine Learning with Amazon SageMaker

Amazon SageMaker is a fully managed service that provides every developer and data scientist with the ability to build, train, and deploy machine learning models quickly. SageMaker removes the heavy lifting from each step of the machine learning process to make it easier to develop high-quality models.

ABOUT THE ANALYST



Dave Schubmehl is research director for IDC's Artificial Intelligence Software Platforms and Content Analytics research. His research covers information access and artificial intelligence technologies including content analytics, search systems, unstructured information representation, cognitive computing, deep learning, machine learning, unified access to structured and unstructured information, Big Data, visualization, and rich media search in SaaS, cloud, and installed software environments.

IDC Global Headquarters

**5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com**

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2020 IDC. Reproduction without written permission is completely forbidden.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.