Data science and MLOps

Operationalize data science to accelerate AI workflows

Artificial Intelligence (AI) is inarguably one of the greatest technologies today. At the heart of AI is data science, an interdisciplinary field using math, statistics, analytics and scientific methods designed to uncover patterns and build predictions with data, algorithms, machine learning and AI techniques. Organizations are increasingly reliant on data science, but implementation has its challenges:

Data science strategy and planning
A lack of clear, concise goals approved by the business owner and other stakeholders can cause confusion, waste the data scientist's time and result in unusable models.

Data collection
Algorithms and models are only as good as the data used to create them. It can be challenging to manually aggregate fragmented data from multiple internal and external resources, in various formats and from both on premises and in the cloud.

Data pipelines and preparation
As business requirements change, manual processes can easily introduce errors and lead to inconsistencies with prior processing. Similarly, when the CI/CD version control, scheduling and orchestration are done manually, errors can be difficult to diagnose and affect multiple tasks.

Model building and deployment
Manual model building can be time consuming, making data scientists less productive. Stand-alone tools can help automate some of the tasks but do not cover the entire model lifecycle of creation, monitoring and deployment.

Model monitoring, tracking and retraining
Manual tracking, recording and correcting for model degradation, drift and bias can lead to the lack of transparent, explainable results necessary to meet today's ethical risk and regulatory concerns.

Role of MLOps and data science

Machine Learning Operations (MLOps) is a set of capabilities and processes designed to operationalize data science across the AI lifecycle. MLOps capabilities include:

1. **Automated low code and no code tools**. Automate data science tools across the AI lifecycle streamline and hasten the process of preparing data, building, training and deploying models at scale.

2. **Automated processes**. Provide a complete view of quality data that's governed, self-served and ready for access using permission-based access. Having this single point of access improves security and governance. Build and move models into production quicker and use automated processes to predict model bias and drift for retraining.

3. **Stakeholder collaboration**. As AI stakeholders continue to grow, automated workflows and communication channels can better connect those who need to be informed and those who need to approve.

TechTarget reported that 83% of organizations have increased their AI budgets, and the average number of data scientists employed has grown by 76%. Yet the time required to deploy a model is also going up, with 64% of organizations taking a month or longer.[1]

IBM data science and MLOps

Built on IBM® Cloud Pak for Data, IBM data science & MLOps provides an end-to-end data science platform automating data collection, model building, deployment, monitoring and governance. The following list includes featured components:

Integrated visual tooling
Prepare data quickly and develop models visually with IBM SPSS Modeler in Watson® Studio.

AutoAI
Apply various algorithms, or estimators, to analyze and prepare raw data for machine learning. Automatically detect and categorize features based on data type, such as categorical or numerical.

Model training and development
Build experiments quickly and enhance training by optimizing pipelines and identifying the right combination of data.

Extensive open-source frameworks
Bring your model of choice to production and track and retrain models using production feedback.

Embedded decision optimization
Combine predictive and prescriptive models, use predictions to optimize decisions, and create and edit models in Python, in OPL[2] or with natural language.

Model management and monitoring
Monitor quality, fairness and drift metrics, select and configure deployment for model insights, and customize model monitors and metrics.

Model management
Compare and evaluate models, evaluate and select models with new data, and examine the key model metrics side-by-side.

AI governance
Trace and document the origin of data, models, associated metadata and pipelines at scale. Also drive responsible, transparent and explainable AI workflows, mitigating risk and protecting against growing AI regulations. Learn more

Contact the IBM Expert Labs team to learn how they can help you deliver AI solutions at speed and scale across all stages of the AI lifecycle.

1. "Addressing 3 infrastructure issues that challenge AI adoption", TechTarget, May 14, 2021, 02 Magic Quadrant for Data Integration Tools, Gartner
2. Optimization Programming Language