WHITEPAPER

Successfully Rolling Out Enterprise-wide ML Ops

Authored by
David Raj Daniel
Distinguished Architect - Data Engineering

ALTIMETRIK®

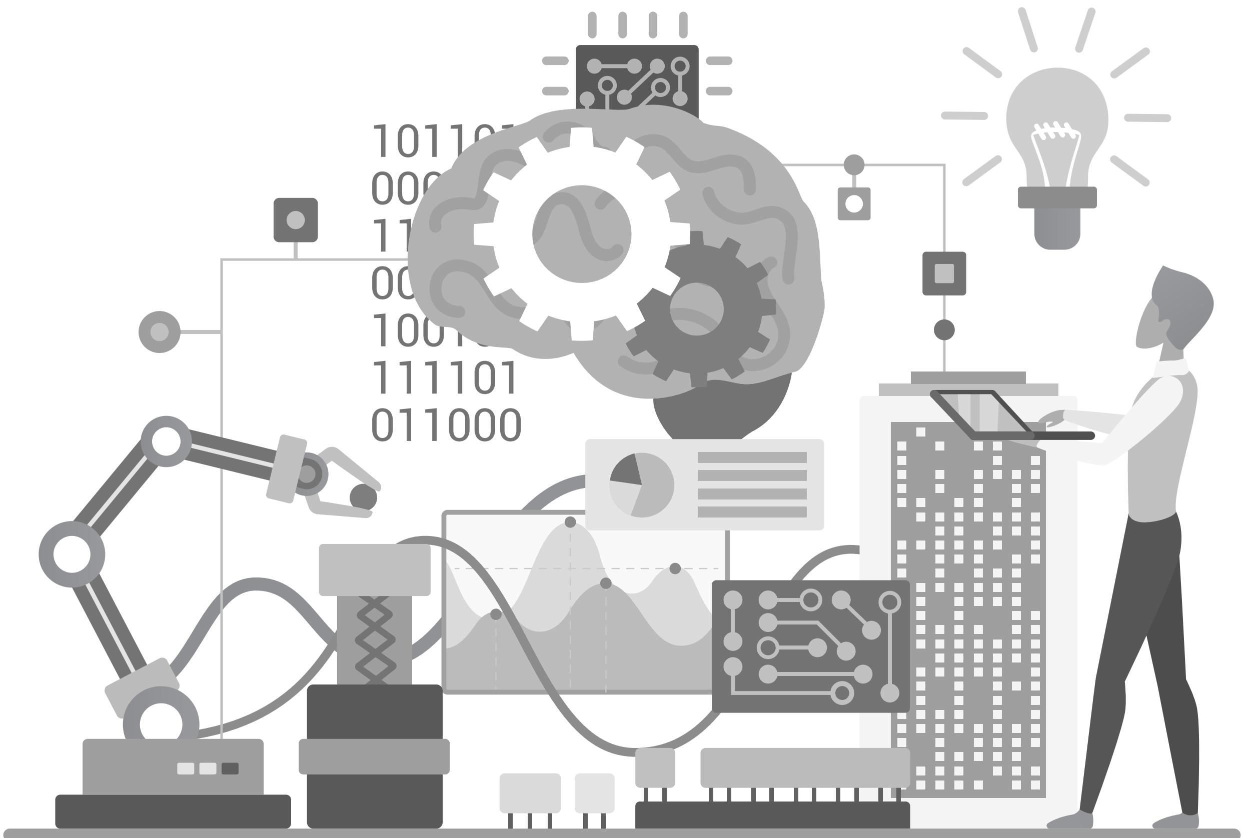
Contents

Introduction

Organizations are becoming adept at launching projects that test their ability to use data, analytics, Machine Learning (ML), and Artificial Intelligence (AI). Data scientists tinker with data sets and analytical models, providing their organizations with the ability to understand trends, test decisions, identify new opportunities, sharpen marketing programs and shape recruitment strategies. These highly-trained data science teams can build sophisticated systems. They can identify missing data values. And they know when their models are going awry. However, data scientists often fail when rolling out and propagating their systems for use by teams across the organization.

The failure can be attributed to several reasons. For example, a home insurance organization's data science team may be using property prices that are not relevant anymore. In production, this model will fail because the data sets required are different. Further, the data used in the lab may be limited. The model may become difficult to scale or degrade with time in real-life applications. Or an organization may feel the process of organization-wide adoption involves multiple teams, which can become challenging to manage. Every large organization has experienced the pain of moving projects from data labs into practical enterprise environments. Before the transition to enterprise-wide usage, there are many challenges to overcome.

To successfully productionize and roll out stable enterprise-wide MLOps, an organization should establish standards. These standards could include the data infrastructure required for the ML lifecycle, data engineering methodologies, ML model engineering, testing, code library/ scalability, model governance, security, tools for MLOps teams, etc.



Creating the Foundation for MLOps

The key to solving the ML Ops puzzle lay in focusing on five crucial steps. These steps create a strong foundation of MLOps for any organization:

Data Exploration

ML uses data for continuous improvement. Therefore, the quality and volume of the training, testing/ validation, and production data sets are critical to success. Inspecting and selecting the right data is the first step to success. The care put into data exploration and quality ensures faster learning and better outcomes.

Data Ingestion

Preparing data for ML applications across functions should be the goal of every organization that wants to extract target ROI from its ML investments. However, there is no such thing as the perfect set of data. Data needs to be cleaned and enhanced; gaps in data must be identified and bridged; and intelligent decisions around which data sets to delete--rather than use--must be taken. More data may not always be the answer to problems (data takes time and money to acquire, store and manage). Some decisions around data engineering can also ensure the asset spans more years of use. Organizations should make the decisions around data ingestions wisely so that they can be easily used across the organization, stay relevant over larger spans of time and business cycles, and keep budgets in control.

Feature Engineering

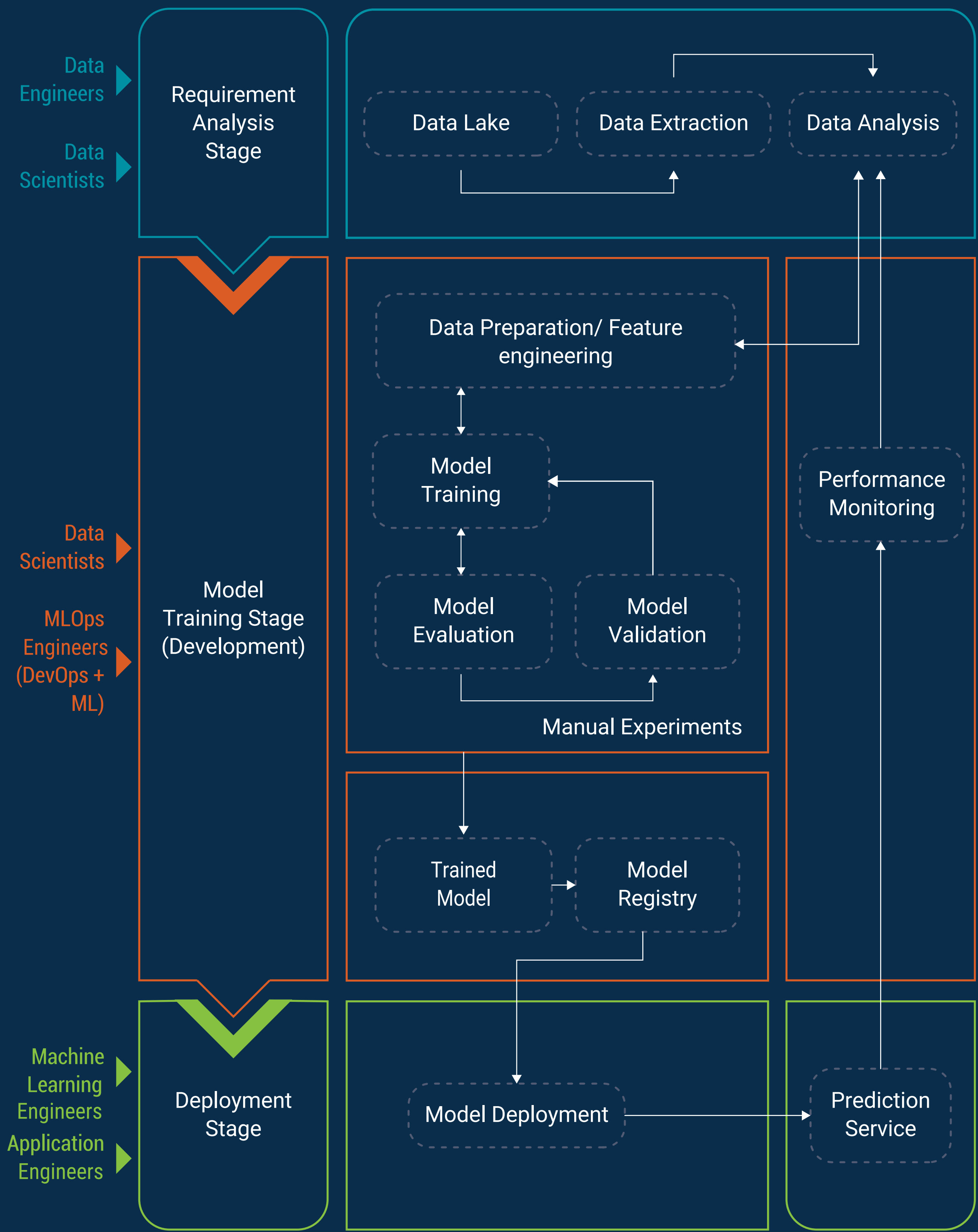
This is a critical step in deriving the features required for model training. MLOps can have a feature store to optimize this process.

Model Training & Tuning

This is a resource-intensive process where ML algorithms and hyperparameters create a model that fits the data. The model is continuously trained and tuned to achieve the desired model metrics using best practices and automating the hyperparameters. A step to evaluate and choose the right model can be added for business use cases.

Model Deployment & Management

Once a model is selected, it is moved seamlessly to production by the MLOps process. Thereafter, model artifacts are managed, model lineage is tracked, and model versions are monitored to keep tabs on data and concept drift. Once in production, a centralized Model Registry is used for two essential tasks. First, to provide a Central Repository enabling teams to collaborate while aiding in managing multiple model artifacts and governance. The second critical task it performs is storing model versions and lineage (when a model was trained, what data was used, algorithms and parameters) and information related to configuration. The model lineage is automatically updated each time a new version of a model is launched. The Model Registry notifies users of every event in the ML lifecycle.



The diagram above puts the five pieces of the MLOps puzzle into perspective.

MLOps Journey of a Pharma Major

A pharma client facing this problem approached Altimetrik with a question familiar to most organizations, “How do we operationalize ML management so that data and analytics become reproducible and reliable across the organization?” What they were discovering was simple: Widely leveraging an ML model in the real world is not the same as building it; they were finding out that while business use cases for ML are growing, they were unable to extract as much value from ML as they should.

The client’s problem could be boiled down to the fact that their ML lifecycle needed to be more efficient, and they needed to migrate their ML systems from on-premise to cloud so they could scale and leverage best practices to be future-ready. MLOps practices are easier to implement using cloud services that providers like AWS and GCP offer. Once the experimentation is complete and the model is being moved into operations, the enterprise will require tools to automate model training, testing, evaluation, deployment, and, monitoring (see Figure 1) which are now available as part of cloud infrastructure.

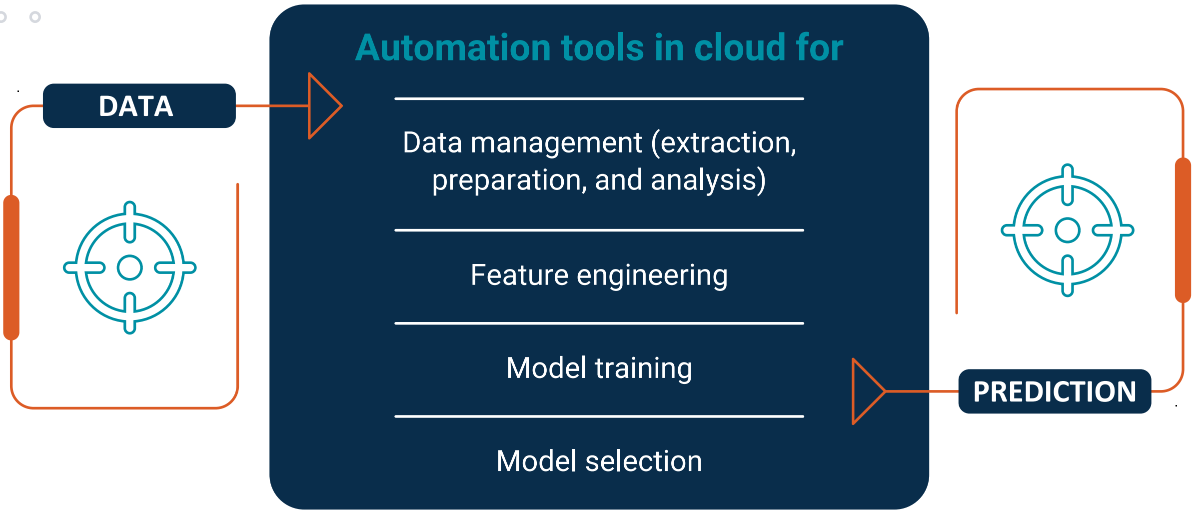


Figure 1

Analysts have found that **88% OF AI/ ML PROJECTS NEVER PROGRESS BEYOND EXPERIMENTATION.** On the other hand, investments in ML programs are growing. In the light of the growing investments, the need to make ML lifecycles efficient is more urgent than ever before. Organizations able to do this will maintain a competitive edge—much like Netflix does by using ML to make accurate suggestions to customers or the way Salesforce does to scale marketing.

Making ML Investments count

Introducing structure to MLOps has aided our pharma client in triggering mass adoption (scaling) and powering an efficient CI/ CD pipeline. ML is now a fundamental part of the client's development process. They have moved their ML investments beyond lab experiments and have accelerated the launch of new features and solutions.

Organizations can gain several benefits from MLOps, just like our pharma client:

Performance gains

By automating the model, the client's process-train-predict cycle has been improved significantly

Cost Control

By modularizing the model development lifecycle to pre-process, train and predict, the client can allocate resources to each step without over provisioning

Maintainability

Development and production lifecycles have been automated and approved by authorized personnel—thus, the client can control ad-hoc human intervention (or bring humans in the loops as required); MLOps monitors and identifies data drift and degradation of models, helping address issues promptly

Scalability

As the data volume increases, the environment scales on demand

Security

Checks for vulnerabilities are maintained in the lifecycle

Audit & Traceability

All artifacts are tracked and controlled, and models are cataloged and versioned—hence behavior is more deterministic and predictable

MLOps is in its infancy. MLOps adoption can be adjusted at varying levels based on an organization's maturity level. Organizations must identify their challenges and explore how best to begin their MLOps journey. In the months to come, organizations will need to take every step they can to improve production speed. MLOps implementation planned and guided by the right set of experienced technologists help them lead the way effectively.

Author Profile

David Raj Daniel
Distinguished Architect
Data Engineering

Distinguished architect and Digital Transformation leader with over 24 years of industry experience in Database Architecture, Database Administration, Application Security, Performance Tuning, Capacity Planning, Machine Learning, Data Analytics, and Enterprise Integration, among many others. Before Altimetrik worked at top Investment & Retail Banks and Others as Data Architect and Performance specialist. Key interests in Data Science, Machine learning, Enterprise data architecture, and Blockchain. Believes in Industrializing Data science projects for maximum value creation. Was part of the Coursera Mentor community for Deep learning and neural network. Holds several certifications in Data Science, SQL, and no-SQL databases. Personal Interests are Linguistics, Medival History, and Philosophy.

About Altimetrik:
Altimetrik is a digital business enablement company helping clients scale digitalization to accelerate revenue growth without disruption. Our practitioners and agile engineering teams take an end-to-end perspective to create solutions in bite-sized pieces through simplification, innovation, and experimentation. Our digital solutions and products provide clients with the tools to fuel business growth and profitability.

Our services and capabilities are tailored for Digital Enablement. Whether it is to enable storied, iconic enterprises to compete with digital disruptors or to enable bootstrapped startups to rapidly launch innovative products, and everything in between, Altimetrik has the right mix of capabilities and a strong track record of success as a trusted digital engineering and enablement partner.

Connect with us

For further information, clarifications and demo, please write to us at

info@altimetrik.com | www.altimetrik.com