

*** New Feature Release *** “Datatron 3.0” – JupyterHub Integration, Kubernetes Management, Model Autocontainerization...[See the New Features!](#)

Datatron Blog

Stay Current with AI/ML

Blog Category

- All
- Machine Learning

Search



What is a Machine Learning Pipeline?

Today, we look forward to learning about an interesting and blossoming process of Artificial Intelligence, i.e., *Machine Learning Pipelines*.

Before we start, let's try to contemplate the terms for better understanding.

‘Machine Learning’ – this is the red-hot iron about which we all have read dozens of articles, compositions, and more. The inference has always been the same, ‘machine learning is an approach to data analysis that involves building and adapting models, which allow programs to learn through experience.’

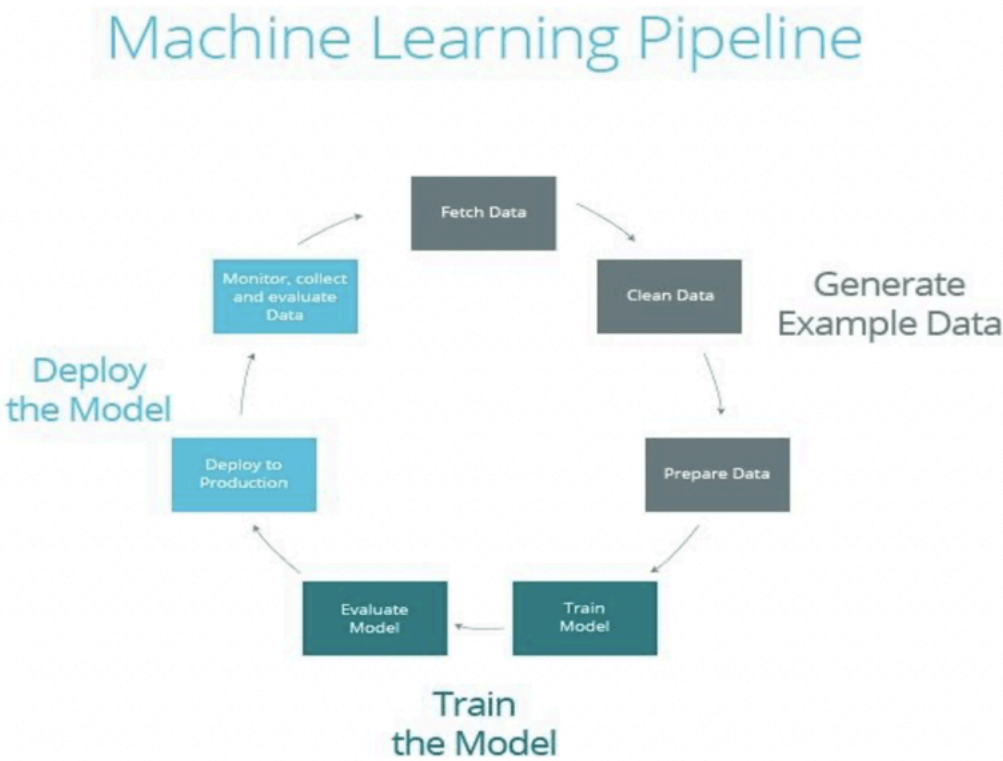
‘Pipeline’ – Have you ever visited a manufacturing plant and seen the assembly lines there where a component progresses through the assembly lines and while passing it is processed on, at different stages simultaneously?

Similarly in computer architecture, **Pipelining** is a technique where commands that are manifold get intersected during execution. The computer pipeline is split into stages. Each stage completes a certain part of a command in parallel. The stages are interconnected to form a pipe in such a way that instructions enter at one end, progress through the stages, and exit at the other end.

Now we will try to integrate the above terms and dive a bit deeper into our today's subject.

What is a Machine Learning Pipeline?

Software developers often talk about ‘build pipelines’ to narrate how software is taken from source code to deployment. Just like developers have a pipeline for code, data scientists have a pipeline for data as it flows through their machine learning solutions. But the term ‘pipeline’ is misleading as it means a one-way flow of data. Generally, a machine learning pipeline encapsulates and automates the multiple Machine Learning processes such as performing data extractions and wrangling, creating training models, model validation, and model deployment for predictions.



As we know MLOps refers to the practice that aims to make machine learning in production efficient and seamless, the Machine Learning pipelines act as the catalyst to this process as it automates the frequent workflow, which in turn accelerates the operation and deployment of the machine learning models. It is easier to fix the bugs in a Machine Learning pipeline that help to create a smoother flow of quality data and reduce the overall cost of development. A well-organized pipeline makes the model implementation more flexible and productive.

Steps involved in a Machine Learning Pipeline

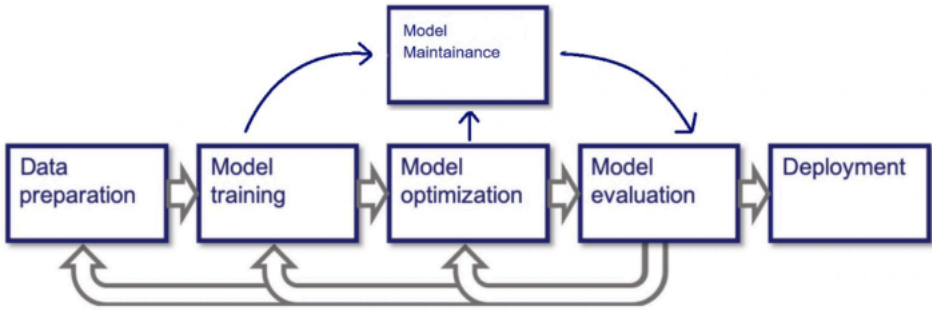
Pipelines are essentially development steps in building and automating the desired output from a program. A machine learning pipeline starts with ingesting new training data and ends with receiving a response on how the recently trained model is performing. The pipeline includes a variety of steps including data processing, model training, and model validation, as well as model deployment and maintenance. One can imagine the fact that going through these steps manually is cumbersome and a very error-prone process.

The machine learning pipeline consists of the following stages:

1. Data preprocessing (exploration and governance)
2. Training models
3. Model Optimization



- 4. Model Evaluation
- 5. Model Deployment



Depending on the specific cases, the final machine learning pipeline may seem to be different. For example, we can train, evaluate, and deploy several models in one pipeline but there are common components that are similar in most of the machine learning pipelines.

Pipeline in Machine Learning for Business Challenges

Following are some of the advantages of using ML pipelines:

- Automates the repetitive processes.
- Easy error identification and fixing.
- Helps to standardize the code.
- Helpful in iterative hyperparameter tuning and cross-validation evaluation.

But what do the above-stated interests offer the businesses? Let us try to answer this question in the section below.

1) Develop more accurate ML models

Using the automated ML pipelines, a smooth flow of quality data can be created that helps the data engineers to tune the ML models that generate more precise outputs.

2) Reach the market expeditiously

Machine Learning pipeline enhances the process of training and fitting ML models so that one can operationalize and deploy them faster, tap into predictions earlier, and get the product to market sooner.

3) Ameliorated business forecasting

ML pipeline helps to build a better model and allows the businesses to strengthen the business forecasting abilities. Improved sales and demand forecasting allow them to stay ahead of the trends, provide a better user experience, and increase profits.

4) Reduces risk

Regular improvements in the ML models through the ML pipeline help to spot risk and opportunities earlier, analyze possibilities more thoroughly, and improve the business strategies.

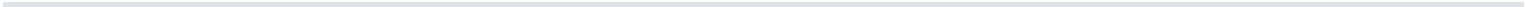
Summing up

Real-life ML caseloads customarily require processes way beyond training and prediction. Data often demands to be processed, sometimes in more than one step. Thus, data scientists must train and deploy not just one algorithm but rather a series of algorithms that synergize to deliver robust predictions from raw data. So, after a glance at the composition above, we can infer that Pipelines in Machine Learning can assist the MLOps engineer to deploy the ML models efficiently and effectively for highly accurate outputs.

Datatron works hard to enable businesses to deploy ML models efficiently so that their customers can focus on tackling the organization's most pressing problems with data-driven insights and decisions.

Follow us on [Twitter](#) and [LinkedIn](#).

Thanks for reading!



BLOG

BLOG

What is a Support Vector Machine?
Datatron Technologies

[previous](#)

Machine Learning For The Modern Enterprise:
The Way Forward For AI ModelOps
Datatron Technologies

[next](#)

WHITEPAPER

5 Reasons Your AI/ML Models are Stuck in the Lab



AI/ML Executive need more ROI from AI/ML? Data Scientist want to get more models into production? ML DevOps Engineer/IT want an easier way to manage multiple models. Learn how enterprises with mature AI/ML programs overcome obstacles to operationalize more models with greater ease and less manpower.



GET WHITEPAPER

view all

Platform

- ModelOps
- AI Monitoring & AI Governance
- Enterprise-ready

Solutions

- Business Executive
- Data Scientist
- Engineering & DevOps

Learn

- Events
- Webinars
- Whitepapers

Company

- About Datatron
- Careers
- Blog
- Privacy Policy
- Terms of Service

Enterprise-Grade,
Production Proven.
Reliable AI™

The Reliable AI™ platform built for enterprise

Built with Love in San Francisco, CA USA!

