

Objective: The objective of this machine learning model deployment project is to develop and deploy a predictive model for a specific use case. The model will be trained on a carefully selected dataset and deployed to make real-time predictions. The project aims to provide a seamless and efficient solution for the targeted prediction problem.

Design Thinking Process: The project followed a design thinking approach, involving empathizing with end-users to understand their needs, defining the problem, ideating potential solutions, prototyping the model, and finally testing and deploying it. This iterative process ensured that the final solution aligns with the users' requirements effectively.

Step 1. Understand the business problem and define success criteria

The first phase of any machine learning project is developing an understanding of the business requirements: You need to know what problem you're trying to solve before attempting to solve it.

To start, work with the project owner to establish the project's objectives and requirements. The goal is to convert this knowledge into a suitable problem definition for the machine learning project and devise a preliminary plan to achieve the project's objectives.

Setting specific, quantifiable goals will help you realize [measurable ROI from your machine learning project](#), rather than implementing a proof of concept that will be tossed aside later.

Is your machine learning project a go or no-go?

Use this chart to gauge the feasibility of your AI project from a business, data and implementation standpoint.

Business feasibility	Data feasibility	Implementation feasibility
<p>Is there a clear problem definition?</p> <p>Is the organization willing to invest and change?</p> <p>Is there sufficient ROI or impact?</p>	<p>Do you have the required data that measures what you care about?</p> <p>Is there a sufficient quantity of data needed to train systems and do you have access to that data?</p> <p>Is the data of sufficient quality?</p>	<p>Do you have the required technology and skills?</p> <p>Can you execute the model as required in a timely manner?</p> <p>Does it make sense to use the model where you plan to use it?</p>

FORBES: QUANTITATIVE, BASED ON AN INTEL CORP. CASE STUDY PRESENTATION
WAVE TECH MARKET, ALL RIGHTS RESERVED

To start a machine learning project, first determine the feasibility of the effort from a business, data and implementation standpoint.

These goals should relate to the business objectives, not just machine learning. Although you can include typical machine learning metrics such as precision, accuracy, recall and mean squared error, it's essential to prioritize specific, business-relevant [KPIs](#).

Step 2. Understand and identify data needs

After establishing the business case for your machine learning project, the next step is to determine what data is necessary to build the model. Machine learning models generalize from their training data, applying the knowledge acquired in the training process to new data to make predictions.

A lack of data will prevent you from building the model, but access to data alone isn't enough: Useful data must be clean, relevant and well structured. To identify your data needs and determine whether the data is in proper shape for model ingestion, focus on data identification, initial collection, requirements, quality identification, insights and aspects worth further investigation.

Even for those with experience in machine learning, building an AI model can be complex, requiring diligence, experimentation and creativity.

But at a high level, the process of designing, deploying and managing a [machine learning](#) model typically follows a general pattern. By learning about and following these steps, you'll develop a better understanding of the [model-building process](#) and best practices for guiding your project.

The right approach starts with identifying data needs and results in a reliable, maintainable final model. In between, you'll work through the stages of data discovery and cleaning, followed by model training, building and iteration.

Step 1. Understand the business problem and define success criteria

The first phase of any machine learning project is developing an understanding of the business requirements: You need to know what problem you're trying to solve before attempting to solve it.

To start, work with the project owner to establish the project's objectives and requirements. The goal is to convert this knowledge into a suitable problem definition for the machine learning project and devise a preliminary plan to achieve the project's objectives.

THIS ARTICLE IS PART OF

[What is machine learning and how does it work? In-depth guide](#)

- Which also includes:
 - [4 types of learning in machine learning explained](#)
 - [How to build a machine learning model in 7 steps](#)
 - [CNN vs. RNN: How are they different?](#)

Key questions to answer include the following:

- What's the business objective and which parts of achieving that goal require a machine learning approach?
- What is the [heuristic](#) option -- in other words, the quick-and-dirty approach that doesn't require machine learning -- and how much better than the heuristic does the model need to be?
- What type of algorithm is the best fit for the problem at hand -- for example, classification, regression or [clustering](#)?
- Have the relevant teams addressed all the necessary technical, business and deployment issues?
- What are the project's defined success criteria and how will the organization measure the model's benefits?
- How can teams stage the project in [iterative sprints](#)?
- Are there requirements for transparency, explainability or bias reduction?
- What are the ethical considerations?
- What are the acceptable parameters for accuracy, precision and confusion matrix values?
- What are the expected inputs and outputs?

Setting specific, quantifiable goals will help you realize [measurable ROI from your machine learning project](#), rather than implementing a proof of concept that will be tossed aside later.

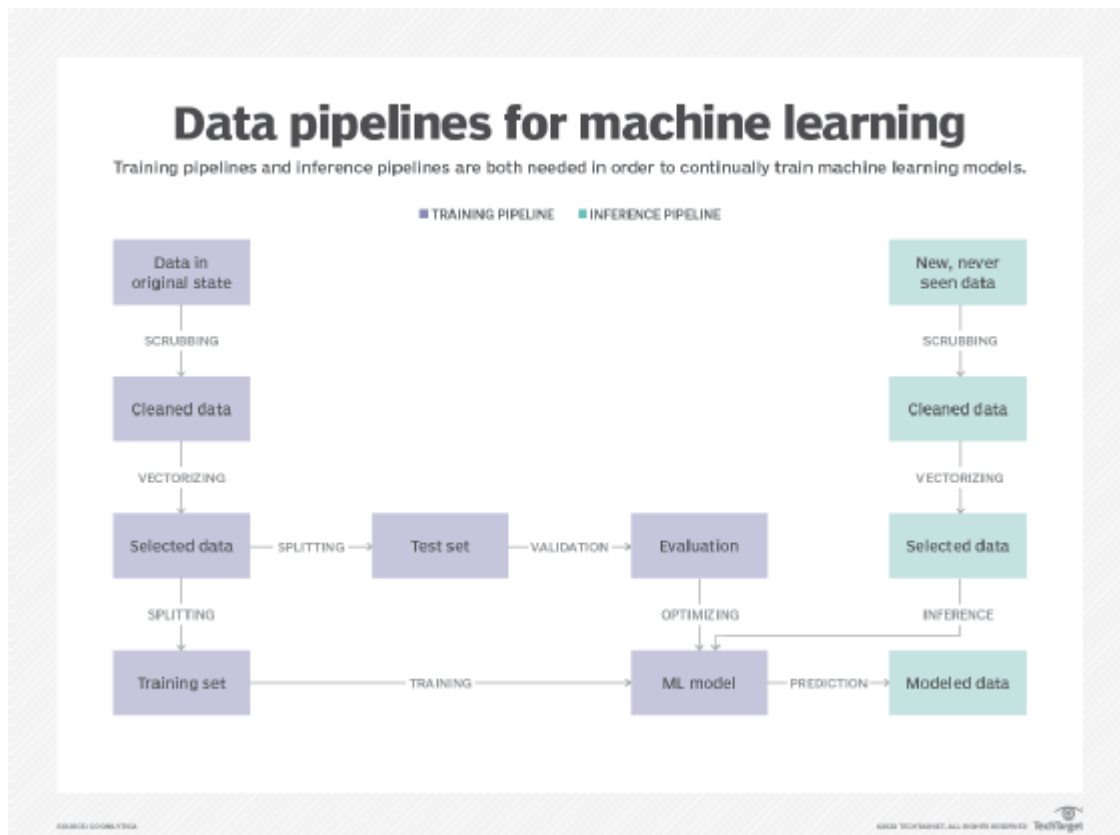
Step 3. Collect, clean and prepare the data for model training

After identifying the appropriate data, the next step is to shape that data so that it can be used to train the model.

[Data preparation tasks](#) include data collection, cleansing, aggregation, augmentation, labeling, normalization and transformation, as well as any other activities for structured, unstructured and semistructured data. Data preparation and cleansing tasks can take a substantial amount of time, but because machine learning models are so dependent on data, it's well worth the effort.

Steps you might undertake during data preparation, collection and cleansing include the following:

- Collect data from various sources.
- Standardize data formats and normalize data across different sources.
- Replace incorrect or missing data.
- Enhance and augment data, possibly by using third-party data or multiplying [image-based data sets](#) if the core data set is insufficient for training.
- Add dimensions with pre-calculated amounts and aggregate information as needed.
- Remove extraneous and redundant information, known as [deduplication](#), and remove irrelevant data.
- Reduce noise and remove ambiguity.
- Consider [anonymizing](#) personal or otherwise sensitive data.
- Sample data from large data sets.
- Select features that identify the most important dimensions and, if necessary, [reduce dimensions](#) using a variety of techniques.
- Split data into training, test and validation sets.



Training and inference pipelines ensure data scientists and engineers can continuously develop and update machine learning models.

Step 4. Determine the model's features and train it

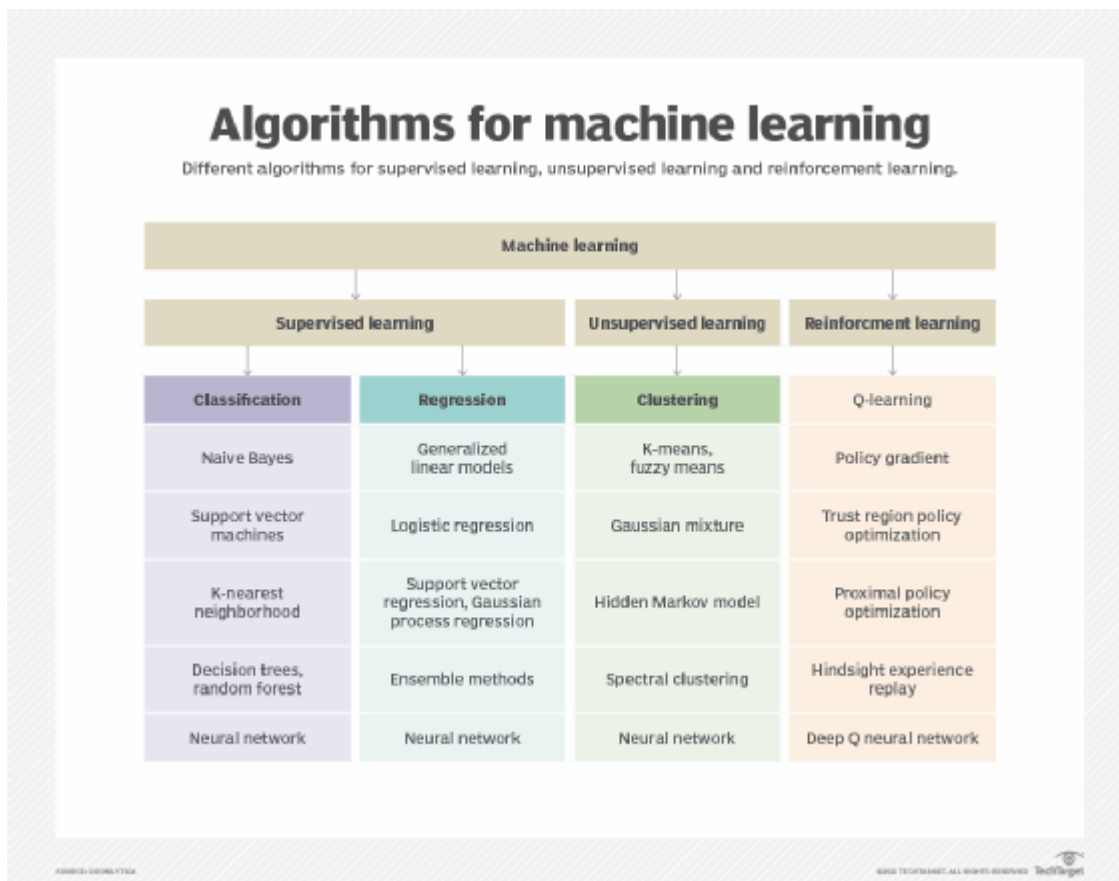
Once the data is in usable shape and you know the problem you're trying to solve, it's time to train the model to learn from the quality data by applying a range of techniques and algorithms. This phase requires selecting and applying model techniques and algorithms; setting and adjusting hyperparameters; training and validating the model; developing and testing [ensemble models](#), if needed; and optimizing the model.

To accomplish all that, this stage often includes the following actions:

- Select the right algorithm for your learning objective and data requirements. For example, [linear regression](#) is a popular option for mapping correlations between two variables in a data set.
- [Configure and tune hyperparameters](#) for optimal performance and determine a method of iteration such as learning rate to attain the best hyperparameters.

- Identify features that provide the best results.
- Determine whether [model explainability](#) or interpretability is required.
- Develop ensemble models for improved performance.
- Compare the performance of different model versions.
- Identify requirements for the model's operation and deployment.

Evaluate the resulting model to determine whether it meets the business and operational requirements.



Machine learning algorithms are formulas or instructions that enable models to improve over time by 'learning' from experience. Different algorithms perform better than others depending on the task at hand.

Step 5. Evaluate the model's performance and establish benchmarks

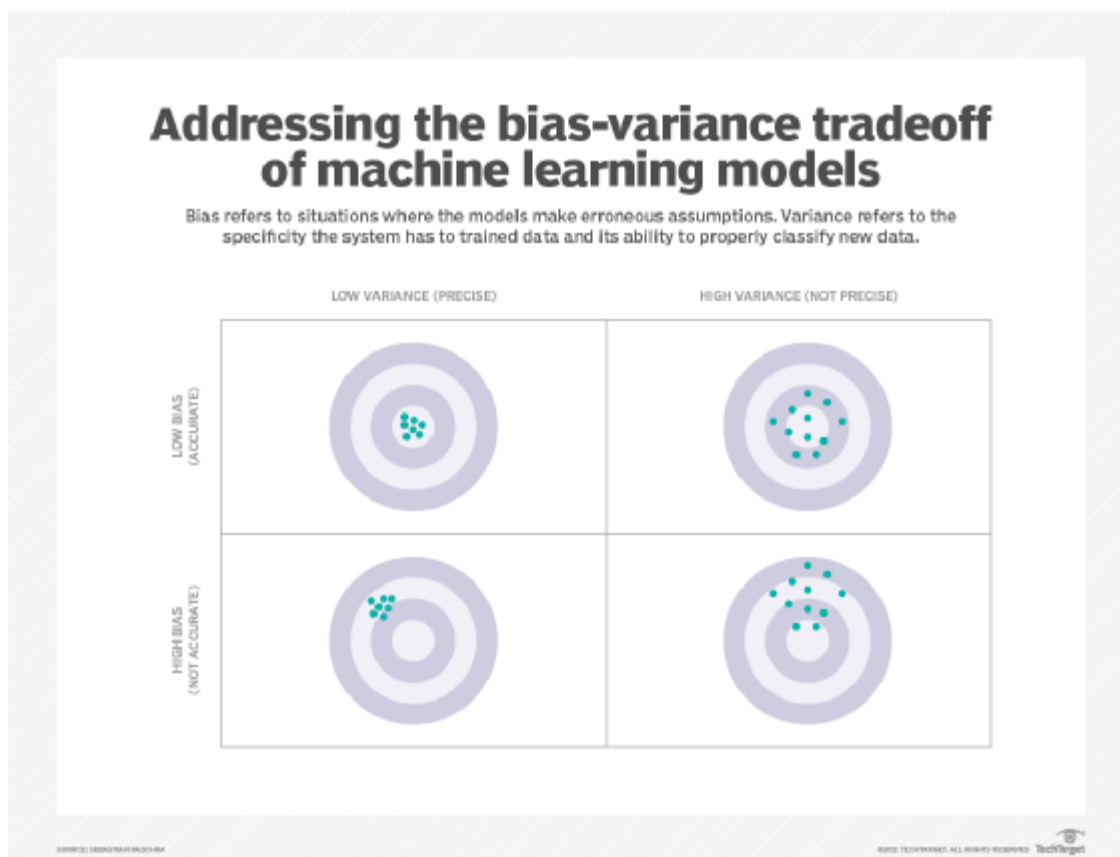
Evaluating a model's performance encompasses confusion matrix calculations, business KPIs, machine learning metrics, model quality measurements and a

final determination of whether the model can meet the established business goals.

During the model evaluation process, perform the following assessments:

- Evaluate the model using a validation data set.
- Determine confusion matrix values for classification problems.
- Identify methods for [K-fold cross-validation](#), if using that approach.
- Further tune hyperparameters for optimal performance.
- Compare the machine learning model to the baseline model or heuristic.

Consider model evaluation to be the quality assurance of machine learning. Adequately evaluating model performance against metrics and requirements helps you understand how the model will work in the real world.



Understanding the concepts of bias and variance helps developers find the sweet spot for optimizing a machine learning model's performance.

Step 6. Deploy the model and monitor its performance in production

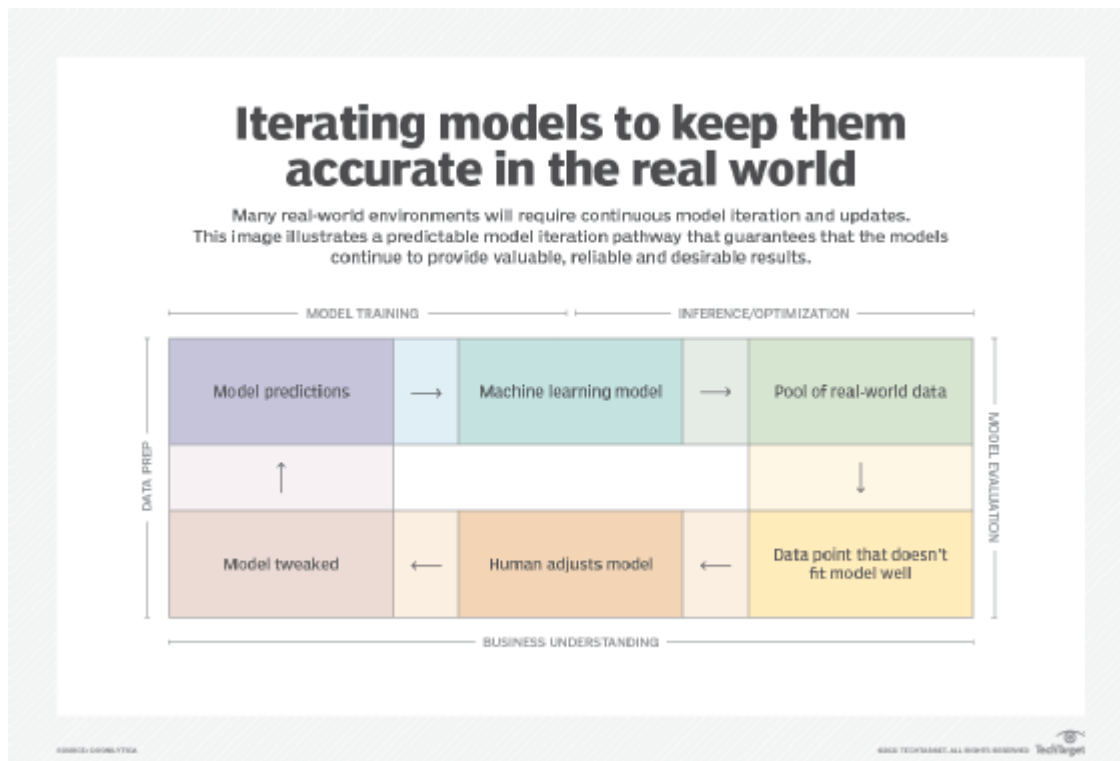
When you're confident that the machine learning model can work in the real world, it's time to see how it actually operates.

This process, known as [operationalizing the model](#), includes the following steps:

- Deploy the model with a means of continually measuring and monitoring its performance.
- Develop a baseline or benchmark against which you can measure future iterations of the model.
- [Continuously iterate](#) on different aspects of the model to improve overall performance.

Operationalization considerations include model versioning, iteration, deployment, monitoring, and staging in development and production environments.

Model operationalization might include [deployment scenarios in a cloud environment](#); at the edge; in an on-premises or closed environment; or within a closed, controlled group. Depending on the requirements, model operationalization can range from generating a report to a more complex, multi-endpoint deployment.



Successful AI projects require iteration to ensure models continue to provide valuable, reliable and desirable results in the real world.

Step 7. Iterate and adjust the model in production

It's often said that the formula for success when implementing technologies is to start small, think big and iterate often.

Even after a machine learning model is in production and you're continuously monitoring its performance, you're not done. Business requirements, technology capabilities and real-world data all change in unexpected ways, potentially creating new requirements for deploying the model onto different endpoints or in new systems.

Repeat the process and make improvements in time for the next iteration. When evaluating and adjusting a [machine learning model in production](#), consider the following:

- Incorporate the next requirements for the model's functionality.
- Expand [model training](#) to encompass greater capabilities.

- Improve model performance and accuracy, including operational performance.
- Determine operational requirements for different deployments.
- Address model or data drift, which can cause changes in performance due to real-world data changes.

Reflect on what has worked in your model, what needs work and what's a work in progress. The surefire way to achieve success when building a machine learning model is to continuously look for improvements and better ways to meet evolving business requirements.