ElectrifAi

How to Accelerate Machine Learning Development

Achieve faster time to value with Ai using strategic investments in people, tools, and data

www.electrifai.net

Introduction

Artificial intelligence (Ai) is starting to deliver results. Businesses have successfully applied machine learning (ML) to achieve positive business outcomes, such as higher accuracy in demand forecasting, faster detection of fraudulent transactions, and predicting a customer's spend passion for upselling.

Yet, 50% of leaders will struggle to move their Ai projects to production, citing a lack of skills, difficulty hiring, and data quality as barriers to entry. The process to build, test, and put a machine learning model into production can take several months.

Companies can accelerate ML development with the strategic deployment of resources at critical stages in the process. Teams can speed the development process by about 65% and increase the likelihood of moving solutions into production by leveraging these resources wisely.

At ElectrifAi, our data science and software engineering teams have built and deployed machine learning solutions for enterprise organizations for 17 years. This white paper shares some of what we have learned about the ML development process over that time, working across industries and use cases.

In this paper, readers will learn:

- The people, data, tools, and time it takes to develop a machine learning model – from proof of concept to production
- Ways teams can accelerate machine learning projects to achieve faster time to value with Ai
- Key factors to consider when teams are preparing to build Ai and put it into production

Contents

CHAPTER ONE The Ai Opportunity for High-ROI Outcomes

Artificial intelligence (Ai) is starting to deliver business results. Organizations use machine learning (ML) to solve challenging problems, from preventing fraud to avoiding customer churn. Across industries, data science and software teams are working together to apply ML to teach machines to detect patterns and signals that would go unnoticed with manual data analysis.

Keep in mind: ML is not the answer to every problem. Organizations can solve business problems with robotic process automation (RPA), business process management (BPM), or data analytics alone. At the same time, others would require too much time or expertise to do manually. There are still many situations in which human decisions provide more accuracy than ML; we recommend that people address them in such cases.

The Power of Machine Learning is Scale

What makes machine learning so powerful is its ability to solve problems at scale. For example, while people are better than machines at extracting information from clauses in a legal contract, it takes weeks for humans to read through thousands of contracts. On the other hand, ML models can complete the same work in seconds, with accurate results – providing the power of massive scale.

In another example, ElectrifAi's team worked with a large healthcare company to use call center transcripts and natural language processing (NLP) to identify common customer pain points and the most effective solutions. Machine learning allowed the team to analyze transcripts from thousands of phone calls to identify similar customer issues, feedback, and problem resolution in a small fraction of the time it would have taken people to do the same work.

The teams used the resulting insights to improve customer experience with more timely interventions to resolve these issues. Other positive outcomes from the engagement included a reduced workload for call center agents and improved training materials to prepare them to resolve these concerns quickly.

Over nearly two decades of machine learning projects, we have learned Ai's return on investment (ROI) can be significant. Companies can generate savings by applying ML to reduce fraud risk, avoid equipment failures, and mitigate customer churn, for example. Outcomes are remarkably positive and significant when the problem to be solved is clear and where data science and engineering resources can work together to examine data sources, identify best-fit tools to use, and collaborate to build models.

High-ROI Outcomes With Ai

Here is an outcome an enterprise organization has achieved using this approach to target specific business challenges:

$2M

saved for an agricultural company by applying a collection of ML models that categorized spend and prioritized savings opportunities.

Similarly, research shows enterprise organizations across industries are applying Ai to achieve high-value outcomes:

- Healthcare: Machine learning modeled potential outcomes of the spread of the COVID-19 virus, as well as the effect of re-opening economies in the wake of the pandemic.[1]
- Oil and Gas: Machine learning powers Shell's precision drilling, which is faster, more accurate, and prevents damage to machinery.[2]
- Manufacturing: Medical device manufacturer Cerapedics uses Ai and predictive maintenance to improve batch yields, lower costs, and mitigate risks.[3]
- Retail: H&M uses ML to select locations for its stores and tailor product availability using local data.[4]

Learn about Ai by use case in our guide:
8 Industries Using Artificial Intelligence to Solve Complex Problems

Enterprise Organizations Are Optimistic About Ai

Even during the COVID-19 global pandemic, leaders of enterprise organizations are optimistic about the promise of Ai, according to a Gartner survey.[5] Nearly half (47%) of Ai investments remain unchanged since the start of the pandemic and 30% of organizations plan to increase their investments in Ai. Only 16% had suspended Ai investments, and 7% decreased investments.

In the same survey, 79% of respondents said their organizations were exploring or piloting Ai projects, yet only 21% said their Ai initiatives were in production. Indeed, developing Ai remains a challenge, with nearly half of leaders struggling to move their Ai projects from proof of concept to production.[6]

In our work with enterprise clients we have learned that the process to conceive, build, test, and put a machine learning model into production can take several months.

According to a Gartner survey

47%
of Ai investments remain unchanged since the start of the pandemic

30%
of organizations plan to increase their investments in Ai

16%
of organizations had suspended Ai investments

7%
of organizations decreased Ai investments

79%
of respondents said their organizations were exploring or piloting Ai projects

7%
of organizations said their Ai initiatives were in production

CHAPTER TWO The Process: Architecting and Building Ai

Let's take a closer look at the Ai development process. We have found this process – from identifying the problem to production at scale – can take six to nine months for the average company leveraging capabilities with a combination of in-house, open-source, and third-party tools and teams.

After production, ML models must be monitored regularly for drift, with recalibrations in training and tuning as needed, to account for changing conditions in the real world, as reflected in data used to operate the ML solution.

Average Time to Build a Machine Learning Model

There are six steps required to architect, build, test, and add machine learning to a system in production:

01

Define the problem, gather data, identify KPIs

02

Identify use cases

03

Explore techniques and solutions

04

Successful proof of concept

05

Train, validate, test model

06

Production

Average Time to Build a Machine Learning Model

01 Define the problem, gather data, and identify key performance indicators (KPIs) Average time for in-house team: 1 month

In the first stage of ML development, organizations must establish a business understanding of the question to be answered or the problem to be solved with Ai. At this stage teams will:

- Determine the problem to be solved. And ask the following question: What is the desired business objective for this project? If the metrics are unclear, the team must surface the pain points and challenges machine learning will address. Teams that fail to achieve this critical milestone in Ai development will surely fail at building an Ai solution that resolves the target pain point. Too often, companies oversee this step, leading to low performance and undesirable outcomes in later stages of development.

- Identify resources and capabilities for the project. In general, examining the people, tools, and data is the right approach here. Teams must answer critical questions, such as:

  PEOPLE:
  What depth of data science, engineering, and DevOps experience on the project team will execute this project? How much domain knowledge do team members have about the target problem and the industry? And, should any manual data work be required, who is available to do that work?

  TOOLS:
  What tools do we have, including proprietary, open-source, and third-party options? Will the project's DevOps environment (e.g., on-premises, cloud, or hybrid) and MLOps approach (e.g., platform, open-source) be during the proof-of-concept phase? What software is available, and what must be procured for the project? Who will need licenses for access?

  DATA:
  What data is available? What are the data sources and format? What is the data dictionary for the database? How can the data be structured and cleaned? Where should data be stored? Is it necessary to enrich the data with information from another data source? What are the feature engineering requirements?

- Determine whether to build tools and models or buy third-party platforms and tools needed for the project; many Ai development teams opt for a hybrid approach. The hybrid choice allows them to build the components of their technology stack that make sense to develop, maintain, and buy parts that an unbiased observer can manage.

- Define key performance indicators (KPIs) that will be used to measure the success of the project. At this stage, teams would typical consider the following: What are our desired outcomes? How will we measure performance – and will our tools allow us to do that? What are the KPIs that would indicate this project was a success?

Average Time to Build a Machine Learning Model

02 Identify use cases

Average time for in-house team: 1 month

In the second stage of ML model development, teams can map potential use cases to solve the target problem. For example, suppose the problem is to reduce costs while improving customer experience. In that case, it might make sense to develop a customer churn model that can predict the propensity of customers' likelihood to churn.

The organization can use those insights to intervene, improve the customers' experience, and achieve the goal to prevent the loss of those customers and the revenue associated with them. A single project may involve more than one use case and result in a collection of ML models put into production across a system.

There is the requirement for data collection and preparation at this stage – and indeed, across the ML development process and into the product lifecycle. This part of the process involves data engineers who prepare data for analysis and for ML models to consume.

03 Explore techniques and solutions

Average time for in-house team: 2-3 months

At the third stage of ML development, data scientists may use exploratory data analysis to discover patterns manually, identify anomalies, test hypotheses, and check assumptions. To do this, they use statistics and data visualization.

The objective of this exercise is to understand the latest approaches and tools available that could be useful for the engagement. Teams must determine which strategies are the best fit based on the data available and the problem to be solved. Tools may be proprietary to the development team's organization, open-source, or developed by third parties.

This stage may also involve acquiring data sets, experimentation with tools and techniques, and feature engineering. Feature engineering is preparing the input data sets to be compatible with the ML algorithm requirements. Here, the team is transforming the data to consider signals or features that have more predictive value.

For example, a model that predicts student success might include an essay as a raw data attribute. Rather than using the essay as input for the model, we can construct additional data features, such as vocabulary size or usage of the Oxford comma, which would provide the model with more predictive power.

This stage of the process is one of the most labor-intensive phases of ML model development. Exploring techniques and solutions requires experimentation, iteration, and tracking results.

Average Time to Build a Machine Learning Model

04

Establish a successful proof of concept (POC)

Average time for in-house team: 1 month

In the fourth stage, teams must establish a POC. Reaching a POC stage means the team can demonstrate the capability of the ML models or collection of models they have developed, using actual outputs, and show how it can solve the target problem. This is a test to help teams understand if a use case can achieve the desired outcomes for the situation they intend to solve.

At this stage, teams can estimate the potential value of the model to be deployed. For example, the team may estimate that a churn prediction model could be used to identify a percentage of customers likely to churn. If the organization establishes interventions to retain those customers, it will save the organization a predicted amount of money revenue that could potentially be lost.

05

Train, validate, and test the ML model

Average time for in-house team: 1-2 months

Just 53%[1] of teams make it to this stage, the fifth stage of ML development. This stage requires teams to train, validate, and test the model they have developed. Here, it is important to determine if the model will perform accurately and achieve desired outcomes when it operates at scale. At this phase, teams also determine what is required to move it into production.

Every machine learning model must be trained to interpret and classify data, predict the likelihood of outcomes, and depending on the algorithm, take action. Training ML models typically requires massive data. It can be challenging for teams to determine when they have enough data to train a model, but in general, the more data the team has available, the better.

At this stage, teams can use hyperparameter tuning or selecting the optimal hyperparameters for a machine learning algorithm. Hyperparameter values are set before the learning process begins. The key to developing high-performing ML models is well-executed hyperparameter optimization.

Three kinds of datasets can be used to train a model:

- A training dataset is used to train a machine learning model. An algorithm or group uses the data to generate a model, mapping data inputs to data outputs.
- A validation dataset is smaller than the training dataset and can be used to evaluate the performance of models with different hyperparameter values. It can also be used to detect overfitting in the training stage.

- A test dataset is used to evaluate the performance of a ML model after hyperparameter tuning. It can also be helpful to compare how different Ai modeling techniques (e.g., neural networks, natural language processing) perform against one another.

Preparation is required to set up processes for ML in production. This can include data operations, annotation processes, quality control (QC) procedures, model monitoring capabilities, and planning for exception handling, which is required when the machine cannot interpret data or take action, so it is flagged for human review and processing.

06 Place the ML model into production

Average time for in-house team: 1 month

Once a model is in production, teams must continuously repeat many of the tasks completed in the earlier stages of ML development. At this stage, they will continue to collect, prepare and annotate data. They will conduct QC on model outputs and identify tasks that can be automated. Automation must be monitored, and exceptions must be processed.

It's also essential to get feedback from a model in production. Monitoring model accuracy and performance over time can help teams surface issues, such as model decay, bias, data skew, and model drift. These problems should be identified and addressed quickly. Feedback must be incorporated into re-training for the model to interpret and act on new conditions in the real world.

CHAPTER THREE

Accelerating ML Development for Faster Time to Value

Finally, let's explore ways teams can accelerate the machine learning development process to achieve faster time to value with Ai. This can be done by making strategic investments in the three main drivers of ML development:

Accelerating ML with PEOPLE

The humans in the loop can determine the success or failure of an Ai project. Machine learning development is accelerated by hiring or outsourcing people or teams that can assist with:

Domain knowledge:
Data science and engineering teams with domain knowledge about an industry and experience building models for the same use case (e.g., churn mitigation, upsell) can significantly speed development time. These teams have experience with ML feature engineering to solve similar problems, and they can move much faster through the last stage, where teams explore techniques and solutions.

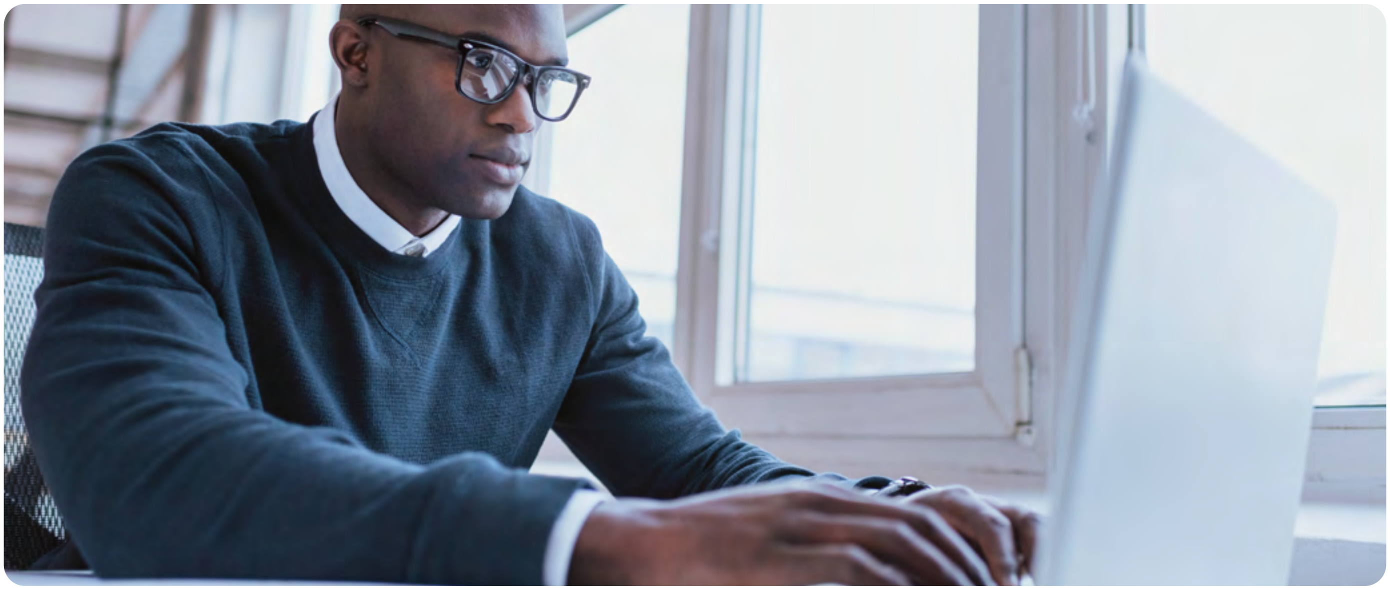
Example:
ElectrifAi Professional Services

Data preparation:

This is sometimes called data annotation or data labeling. ML models are only as good as the data that they are trained on. High-quality data annotation contributes to high-performing ML models.

Automation monitoring:

Later in the ML development process, as teams automate tasks that deliver consistently positive outcomes, people will need to monitor the automation and process exceptions.



Accelerating ML with TOOLS

ML tools and techniques have expanded and improved rapidly over the last decade, giving rise to more effective algorithms, powerful computing capabilities, and platforms to track everything from cloud spend to ML experiments.

This environment will remain dynamic for the foreseeable future. Machine learning development can be accelerated with tools available in one or more of these categories:

DevOps:

Smart infrastructure decisions can shorten the ML development process and ease continuous delivery. Data storage and retrieval are critical components of ML in development and production. ML development can be accelerated and optimized with cloud tools and service providers.

Examples: ElectrifAi, Talend

ML modeling:

Open source and proprietary tools speed algorithm selection, data visualization, hyperparameter optimization, and other critical tasks in the ML development process. It is worth exploring available tools for every ML model developed and upgrading models over time as new tools emerge. Some service providers can manage this part of the process.

Examples: Amazon SageMaker, Dask, ElectrifAi

Platforms:

Enterprise teams are pairing open-source tools with platforms that allow them to spin up ML experiments in the cloud and run them faster, accelerating the runway to production.

Example: Comet ML

Accelerating ML with DATA

Data is the lifeblood of Ai. The choices teams make about data to be used, how it will be processed, and how quality will be managed contribute to ML outcomes.

Acquisition:

In some situations, it is helpful to use free or purchased data. Sometimes, that data is already annotated for a use case, such as image data annotated for training autonomous systems.

Annotation:

The race to usable data – labeled, structured data prepared for a particular use case and model – requires people, training, time, and advanced tools.

Example: iMerit

Enrichment:

Adding to or enhancing data that teams already have can be helpful in some use cases. Feature engineering can reveal opportunities and new approaches to take with existing data, which can accelerate development

Example: ElectrifAi

Getting Started with ElectrifAi for AWS

ElectrifAi uses machine learning algorithms powered by Ai to find patterns in massive amounts of structured or unstructured data. Seamlessly integrating into our client's technology ecosystem, we generate actual business value. Our capabilities and domain knowledge deliver new, pre-structured, and pre-trained machine learning solutions that apply the latest tools and techniques, require minimum resources to implement, and provide positive outcomes with high ROI.

ElectrifAi has redefined the way businesses use Ai and machine learning. Driving revenue uplift, cost reduction as well as profit and performance improvement, we are a global leader in business-ready machine learning models. Our large library of models on AWS Marketplace reaches across business functions, data systems, and teams to drive superior and proven results in record time.

By deploying machine learning solutions that require minimal customization and integrate with existing workflows, our clients gain faster value with Ai and better results – all at a lower cost. Achieve more for your business with ElectrifAi.

Visit ElectrifAi in AWS Marketplace

References

1. Machine Learning Predicts COVID-19 Outcomes, Contagion Live, January 20, 2021
2. Ai Applications in Oil and Gas Industry, SparkCognition, Inc. January 28, 2021
3. Data Quality Key to Leveraging Ai, Machine Learning, 2021. Jeff DeVerter, Manufacturing.net.
4. #Machine Learning in Retail: 5 Juicy Examples | Technologies for Business, Mass Media Group, August 28, 2020.
5. Gartner Survey Reveals 66% of Organizations Increased or Did Not Change Ai Investments Since the Onset of COVID-19, Gartner, October 1, 2020
6. Ai Project Failure Rates Near 50%, But It Doesn't Have to Be That Way, Say Experts, Wall Street Journal, August 7, 2020

Contact us today!

To learn more, visit electrifai.net/partners/amazon-web-services, or contact us at sales@electrifai.net.