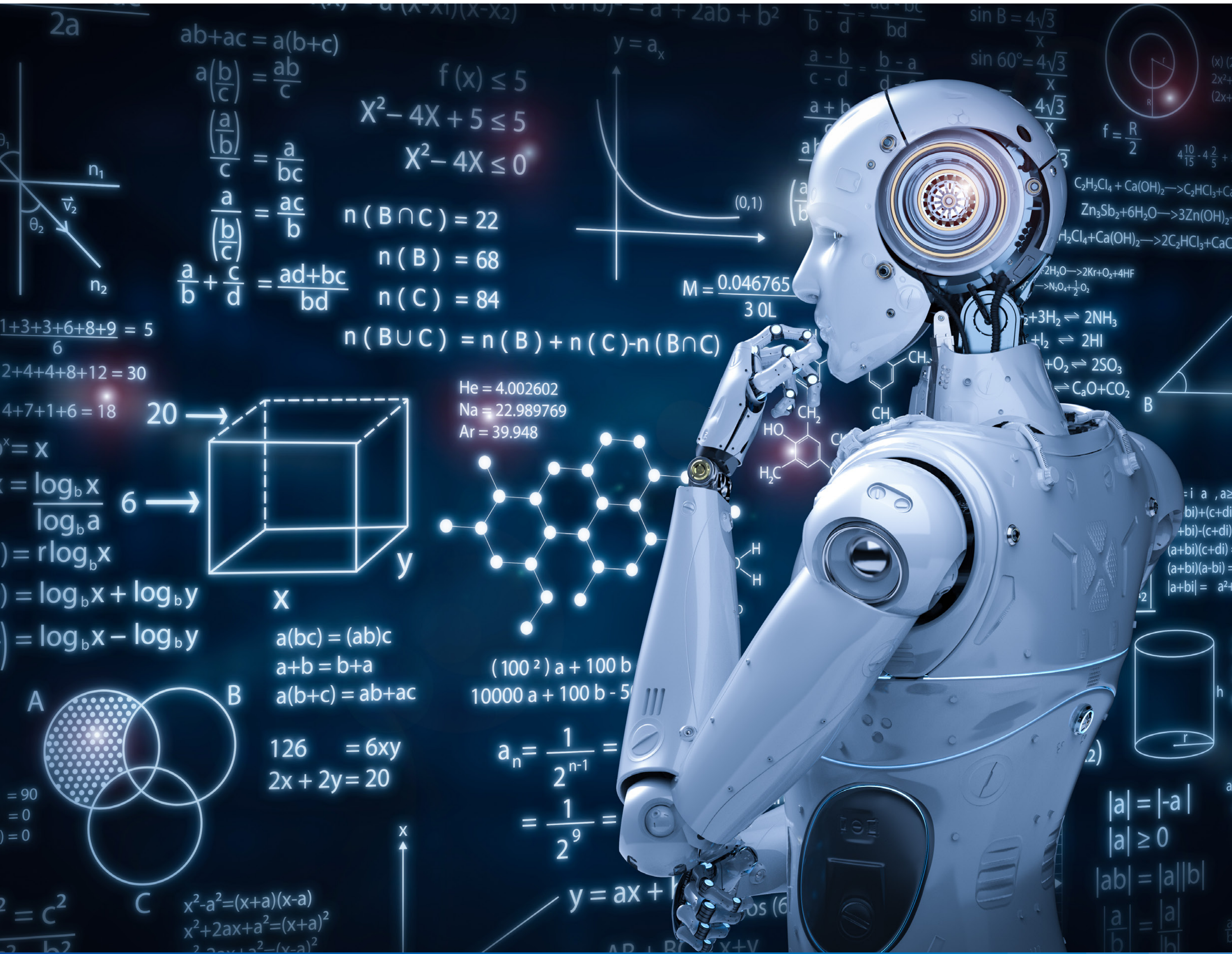
Course5
Transformative intelligence



AUTOMATED MACHINE LEARNING (AUTOML):

An Industry Overview and Implications for Business

Ganesh GJ
Market Research Manager
Course5 Intelligence

Abstract

Enterprises have begun to shift from just solving static business problems to predicting the future. Hence predictive modeling and Machine learning (ML) have become essential technology for enterprises to make data-driven decisions. Major technology companies like Google, Facebook, Amazon, Microsoft, and others have actively reoriented themselves around Artificial Intelligence (AI) and ML using an 'AI-first' strategy. These companies have also invested time and engineering effort on creating their own machine learning-specific tools. Over the last five years, AI research has grown by 12.9% annually worldwide, according to popular technology writer Alice Bonasio.

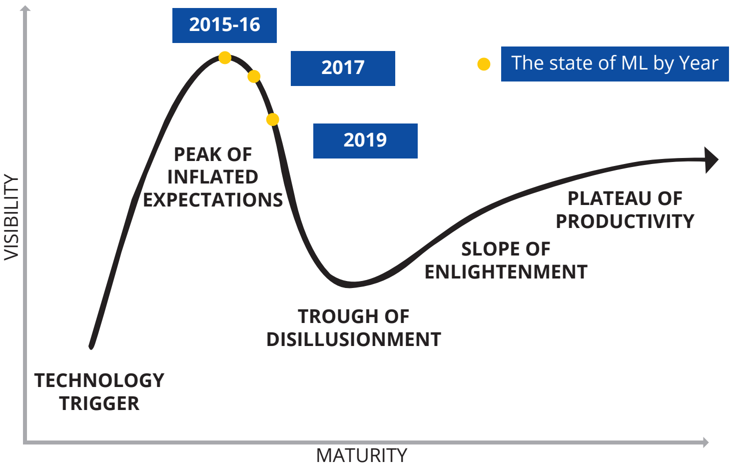
The Current Hurdles in ML Industry

Lack of Skilled Talent: Implementing AI/ML is time-consuming and finding the right talent is even more challenging. There is a growing need for quick turnaround time on AI projects and the demand for skilled data scientists to enable this has increased exponentially. The requirement for data science expertise has outpaced the supply of this skillset despite the surge of people entering the field. Further, the ML models are continuously evolving and the scenarios are getting more complex day by day.

Low Productivity and ROI: Another major challenge facing the ML industry is the need to increase the productivity of data scientists and engineers in building efficient ML applications. Due to the advent of emerging technologies like Big Data, 5G, etc., there is a huge volume of data generated that calls for efficient ML algorithms to analyze this complex data. With the burgeoning flow of data streams, data scientists can no longer keep building ML applications manually, as designing neural nets is going to become extremely time-intensive and will require multiple industries and domain expertise.

Failed ML Efforts: The ML industry is also affected by failed long-term ML projects, not generating the expected ROI, often due to lack of skilled data scientists with lack of sample data for training. Also, the lack of a common connecting platform between developers and domain experts makes it tedious task to connect different kinds of people involved in the ML development process, like data engineers, Business Analysts, DevOps, ML engineers, citizen data scientists and software engineers. The solution lies in an automated approach, which has the ability to bring a fundamental shift in the way organizations approach ML and data science. This requires the right set of automation tools and a common platform that shortens the time-to-value for data science teams and connect the entire ecosystems of the AI & ML development process.

ML has transitioned from a state of hype to being one of the most sought-after technologies across industries



Gartner Hype Cycle for Emerging Technologies 2015-2019

Fig 1: Gartner Hype Cycle for Emerging Technologies 2015-2019

Source: Gartner 2016, 2017, 2019

During the period from 2015 to 2019, Machine Learning moved from the peak of its hyped state to having many tangible use cases, especially in the B2B segment, leading to increased customer trust and belief. Today, ML has become one of the most sought after technologies across all industry verticals and horizontals mainly due to increased availability of large sets of open data, hosts of ready-to-use algorithms and open source programs. AI/ML has become an indispensable part of any organization in their digital transformation journey.

Tech-leaders like Google, Facebook, Microsoft, and Amazon, who have tons of customer data at their disposal, have popularized the use of ML and data science to build their businesses and achieve an edge over the competition. Furthermore, many AI/ML-based startups have mushroomed across the globe with specialization in some of the niche segments. However, the AI/ML industry is still fragmented and there is a huge shortage of talent as well.

Gartner HypeCycle for Artificial Intelligence 2019

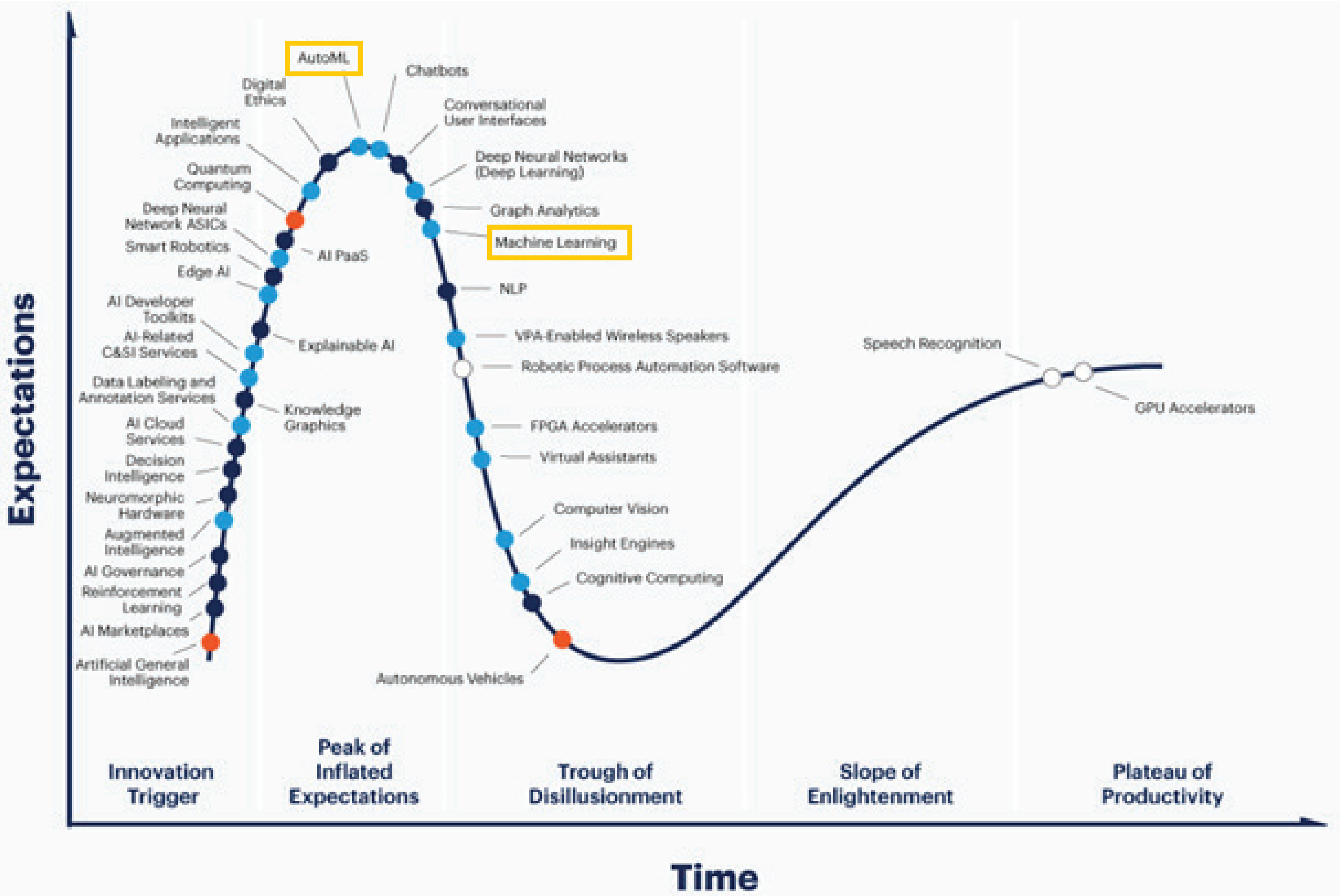


Fig 2: Gartner Hype Cycle for Artificial Intelligence, 2019

Source: Gartner HypeCycle 2019

Due to the widespread adoption of machine learning across the industry, Automated Machine Learning (AutoML) has gained traction across the globe. AutoML is following a path similar to ML and is currently positioned at the peak of inflated expectations in the Gartner hype cycle, a position which was occupied by ML in 2015-16. In the future, it is expected that a number of small players will develop niche AutoML capabilities along with massive investments by technology giants like Google, Amazon, and Microsoft. Due to these combined efforts, AutoML may soon move from a state of hype to having many tangible use cases in the future.

Traditional ML is time-consuming and resource-intensive, requiring significant domain knowledge

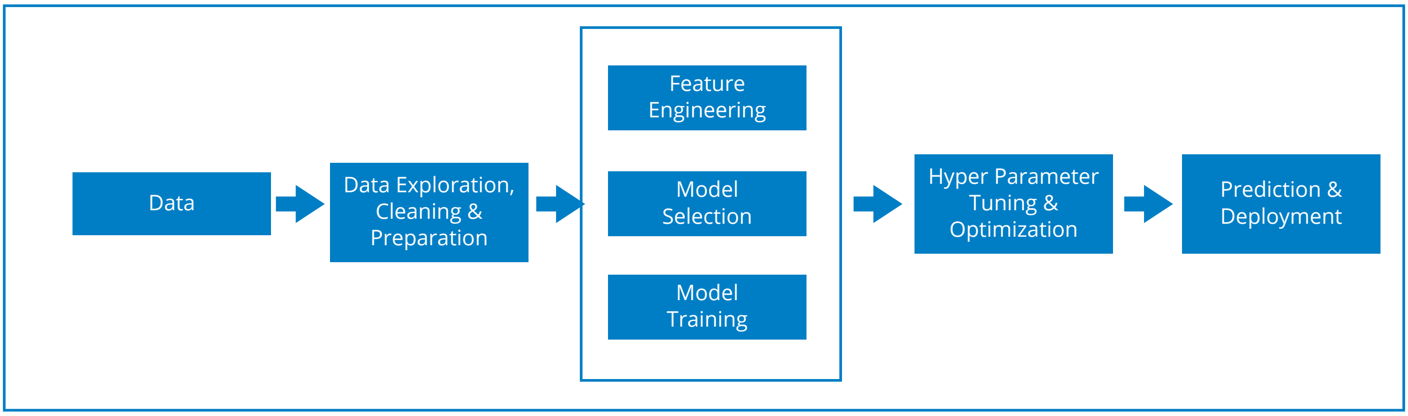


Fig 3: Traditional Machine Learning Workflow

Traditional ML model development is an iterative process that is resource-intensive, requiring significant domain knowledge and time to produce and compare dozens of models. Traditional ML no doubt automates a business process from a customer's point of view, but the process of developing such ML models itself today needs automation due to the advent of connected devices, IoT, Big Data, and other emerging technologies. AutoML addresses this need by enabling data scientists to generalize and automate some of the most difficult tasks like feature engineering, hyperparameter tuning and others involved in building ML solutions. This makes AutoML an extremely extension part of AI/ML practice.

A Growing Workload with Traditional ML Practices

In the traditional ML process, the Data Engineering team generally works on data acquisition, data preparation, and building and optimization of models. The Development Operations team typically focuses on the development environment, tooling, and hosting of the inference models in production. In the current scenario, an expert has to manually check the data type and then apply the appropriate data pre-processing, feature engineering, feature extraction, and feature selection methods that make the dataset amenable for the ML process. Following those pre-processing steps, practitioners must then conduct algorithm selection and hyperparameter optimization to maximize the predictive performance of their final Machine Learning model. Apart from these tasks, practitioners have to manually test, train, tweak and again test the models. Overall, these are extremely time-consuming and cost-intensive tasks involving several teams, with interdependent tasks often beyond the capabilities of non-experts.

Time Spend by Data Scientists

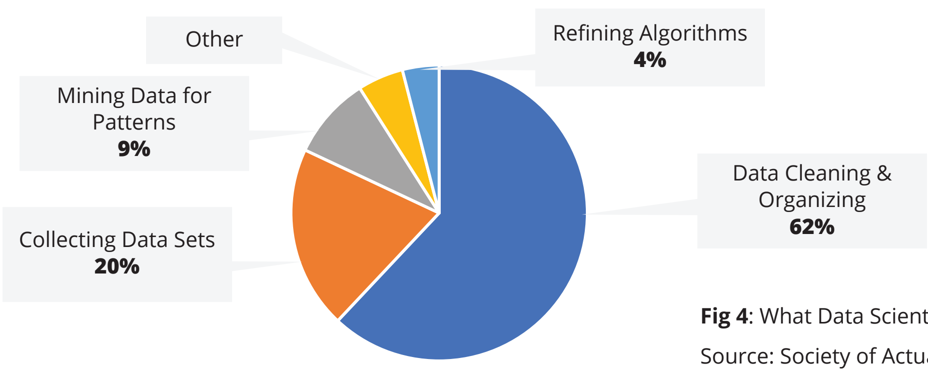


Fig 4: What Data Scientists Spend the Most Time Doing

Source: Society of Actuaries - The Past, Present, and Future of AutoML Report, 2018

60% of the Data Scientist's time is spent on non-core tasks like Data Cleaning and Organizing Data.

What is AutoML?

AutoML refers to automated methods for model selection and/or hyperparameter optimization. It provides an effective and efficient method to generalize and automate solutions to similar problems, and thereby addresses the growing challenge of applying ML faster and more widely to various industry needs. The goal of AutoML is to shorten the cycle of experimentation, trial, and error which involves multiple stages like hyperparameter tuning, model selection, feature engineering, etc. This is a tedious and time-consuming activity for any data scientist in the long run. AutoML platforms can perform these repetitive tasks more quickly and exhaustively to reach a solution faster. This also frees up data scientists' time and allows them to focus their energy and attention on other aspects of the process that require a higher level of thinking and creativity. With AutoML, hundreds of thousands of developers need not design new neural nets for their particular needs. Instead, they can simply leverage the existing AutoML products by making necessary tweaks.

AutoML is creating massive possibilities for marketers, enabling deeper customer engagement on a personal basis. Whether the idea is to empower data scientists/business users or to boost productivity, AutoML solutions are quickly becoming a must-have for every organization looking to scale up their use of ML.

According to Forrester, most organizations can benefit from a standalone AutoML solution, and this market is expected to grow substantially as products get better and there is an increase in awareness on how these tools fit in the broader data science, ML, and AI landscape. Apart from standalone AutoML providers, multiple cloud vendors like AWS, Microsoft Azure and Google Cloud have also launched AutoML services in the recent past, which clearly hints to growing use of AutoML in the future.

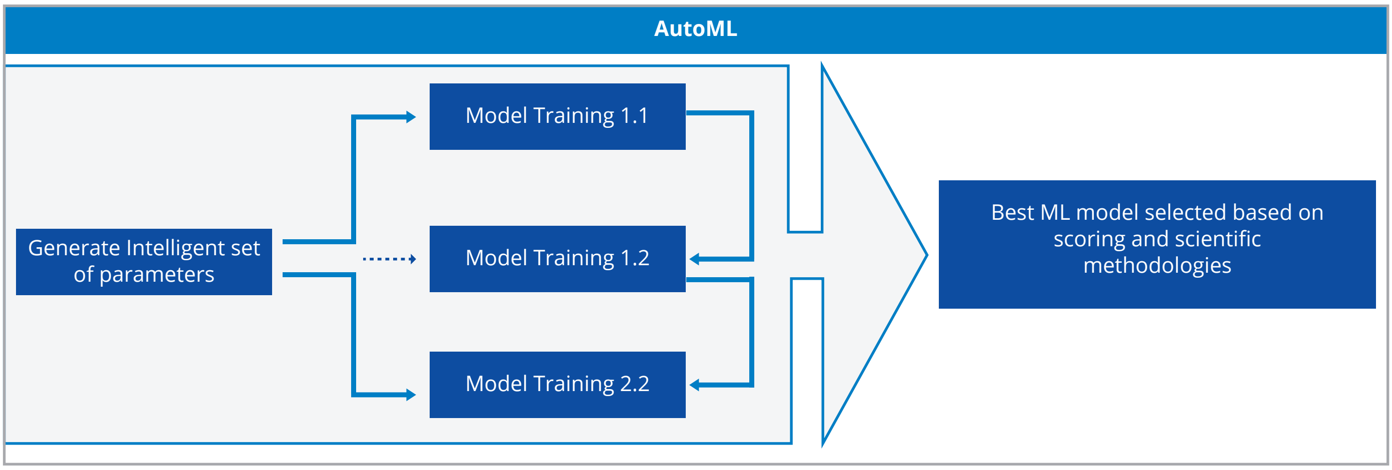


Fig 5: Typical Working Process of AutoML | Source: Activeeon

AutoML accelerates the process of evolving a trained model by automating some of the major steps in the process. It simplifies the process of identifying the best hyperparameters by selecting a combination that will optimize the model training score. This gives every scientist and researcher an easy tool to arrive at the best model within limited resources in terms of computing power, money, and timeFrom feature engineering to hyperparameter tuning, AutoML automates the most complex steps of the pipeline. AutoML creates a new class of 'citizen data scientists' that puts the power of advanced ML directly in the hands of business users.

Further, AutoML also allows the required level of customization without forcing developers to go through the elaborate workflow. In recent years, several off-the-shelf AutoML packages have been developed: One exampleis the Driverless Vehicle where the AI systems are powered to create and train surprisingly good models in a short time, without requiring data science expertise. Many commercial organizations have attempted to automate Machine learning, creating toolkits such as Auto-sklearn, Auto-Weka, Google AutoML, and H2O.ai's Driverless AI, etc.

When to use AutoML?

AutoML empowers data scientists with limited expertise to develop ML models and to identify an end-to-end machine learning pipeline for any problem.

Data scientists, analysts and developers across industries can use AutoML in the following scenarios:

- Lack of highly skilled talent
- Enabling non-experts to train high-quality machine learning models
- Need for rapid prototyping and testing
- Repetitive tasks that exhibit similar expressions
- Reducing time in model development and resources
- Leveraging data science best practices
- Providing agile problem-solving

Why AutoML has gained traction?

- Widespread adoption of Machine Learning across the industry
- Limited availability of skilled data scientists and ML experts
- The need to find insights from a huge set of multi-dimensional raw data
- The need for automation for time and cost savings
- Increasing availability of AutoML libraries and tools and consequent awareness of AutoML's potential

Key stages in the ML pipeline where automation can be implemented

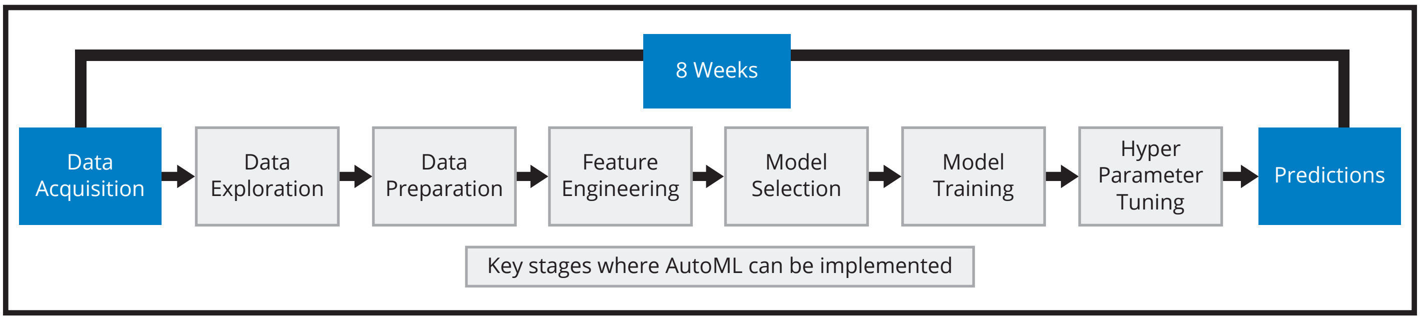


Fig 6: Machine Learning Pipeline

To implement ML traditionally, depending on the volume of data, it would approximately take 8 weeks. The stages from data exploration to Hyperparameter tuning may be automated using AutoML methodologies which will help reduce this time. For some stages, full automation may not be possible due to the nature of data and unique business problems.

Key stages in the Auto ML Pipeline

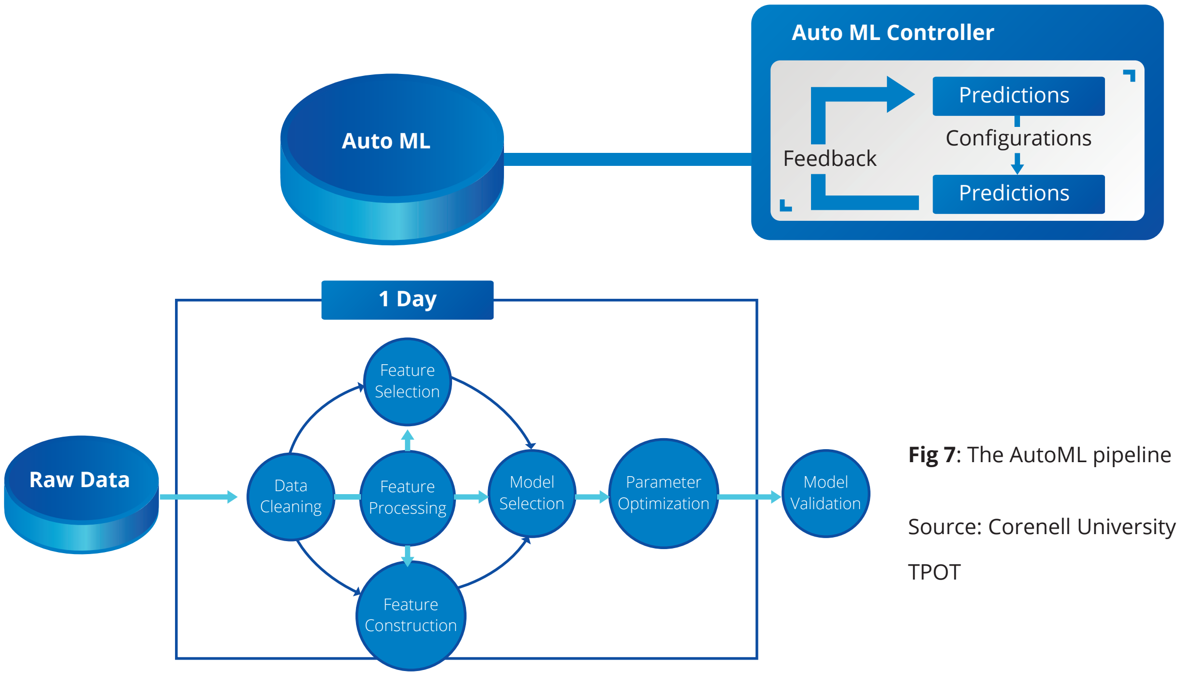


Fig 7: The AutoML pipeline

Source: Corenell University study on AutoML 2019; TPOT

The implementation of AutoML in typical ML practices helps reduce turnaround time from approximately 8 weeks to just 1 day and further frees data scientists' time to focus on other important activities like data-preprocessing, model deployment, etc.

Comparison of Traditional ML and AutoML – What can be automated?

	Traditional ML (Human-driven, manual)	AutoML
Feature Engineering	Feature design and selection	Automated by the computer program (AutoML)
	Reiterative process of making features more informative, Data Binning and transformation	
Model Selection, Validation and Assembling	Algorithm Selection, Ensemble and manage other ML tools	
	Adjusting hyperparameters of ML tools based on performance evaluation, Retraining	
Algorithm Selection	Selecting optimization algorithms to find parameters	
Summary	Human experts are involved in every aspect of ML application	The program can be directly re-used on other learning problems

Table 1

Key sub-stages that need to be automated to implement AutoML

What can be automated	How it can be automated
Data Preparation and Ingestion (Preprocessing)	• Automated pre-processing steps that include feature normalization, handling missing data, converting text to numeric data, etc. • Automated column type detection; e.g., boolean, discrete numerical, continuous numerical, or text • Automated column intent detection; e.g. target/label, stratification field, numerical feature, categorical text or free text feature • Automated task detection; e.g., binary classification, regression, clustering, or ranking
Feature Engineering	• Automated feature creation from multiple related tables using feature tools and frameworks that use transformations and aggregations. Also use Python libraries like Pandas. • Automated Feature Selection • Automated Feature Extraction • Meta-learning and Transfer Learning • Detection and handling of skewed data and/or missing values
Model Selection	• Search Space: Grid Search, Random search, etc. • Classification tools: Once features have been obtained, the program automatically finds a model like SVM, deep networks, logistic regression, decision tree, random forest, etc. to predict the labels

Hyperparameter Optimization	• Blackbox Hyperparameter Optimization
	  ○ Model-Free Blackbox Optimization Method
	  ○ Bayesian Optimization
	• Multi-Fidelity Optimization
	  ○ Learning Curve-based Prediction for Early Stopping
	  ○ Bandit-based Algorithm Selection Methods

Validation & Assembling	• Automatically chooses adequate cross-validation techniques, evaluates variable contribution, and protects against data leaks • Automatically combines models to obtain optimal models, and re-tune them to get optimal, stable output

Table 2

Industry applications of AutoML

Healthcare

The healthcare industry requires accuracy and highly interpretable production-ready models a short time. Busy clinicians and health care researchers are in need of supportive tools and computing resources in developing models that enable fast turnaround.

Major use cases of AutoML in Healthcare:

- Clinicians without prior experience in coding or deep learning are able to develop models to accurately detect common diseases from medical images using AutoML
- Quick and accurate diagnosis through analysis, classification, and labeling of medical images and creation of custom vision models for image recognition
- Healthcare Analytics to predict patient needs
- Obtaining insights from incomplete medical health records
- Clinical research, especially clinical prediction/classification problems
- Creation of patient journeys for specific disease stages and finding probability at each stage

Retail

AutoML can enhance the customer shopping experience with better discovery, recommendation, and search capability. Furthermore, AutoML helps retailers automate the product attribution process by recognizing nuanced product characteristics.

AutoML can help retail businesses identify new customers, manage SCM, set pricing, optimize promotions, predict retail spending and sales, and create better customer experiences.

Financial Services

Changing consumer preferences, increasing competition from financial technology companies and complex regulatory requirements are driving rapid change in the Finance and Banking industry. Equipped with powerful proprietary data, banks and credit unions are planning to use AutoML to make better decisions, offer better customer services, and streamline operations.

AutoML in the financial services industry is still in the exploratory stage where people have found the concept interesting and are actively evaluating platforms. Accelerating project turnaround time, curiosity to find deeper business insights, and the need to democratize AI and ML outcomes are the key drivers of AutoML in the Financial services.

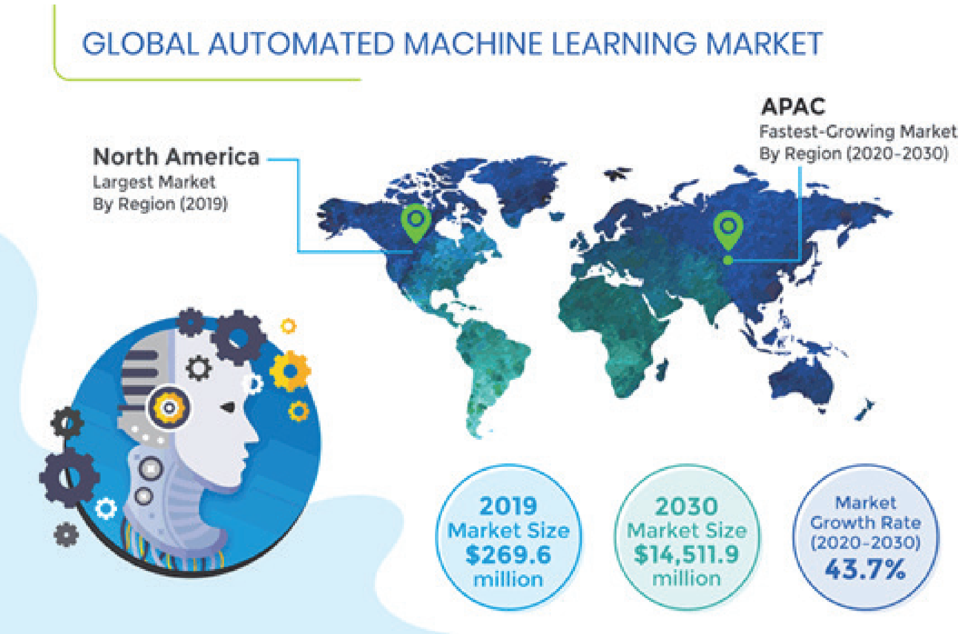
Major use cases of AutoML in Financial Services:

- Loan Repayment Prediction
- Forecasting Client Creditworthiness
- Fraud Prevention
- Predicting Customer Churn

Market Overview of AutoML

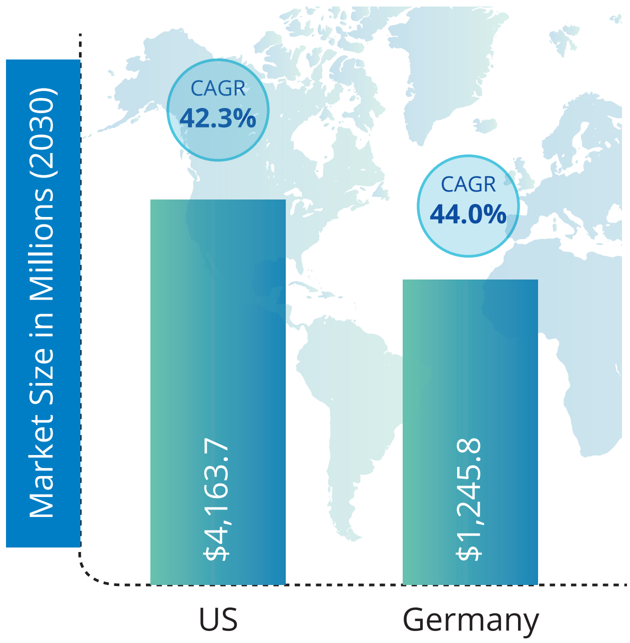
In the past few years, there has been a surge of interest in AutoML tools that automate a range of tasks in the data science workflow and the market is posed to grow rapidly in the next 2–4 years. AutoML currently does not outperform hand-constructed models in terms of accuracy, but can achieve as much as 95% of the performance of the best performing model in most cases.

According to P&S Intelligence market report, the global AutoML market generated revenue of $269.6 Mn in 2019 and is expected to reach $14,511.9 Mn by 2030, growing at a CAGR of 43.7% during the forecast period between 2020–2030. Increasing demand for efficient fraud detection solutions, growing need for personalized product recommendation, and rising importance of predictive lead scoring are some of the key factors driving the growth of the market across the globe. APAC is expected to record the fastest growth during the forecast period. This can be attributed to the rising economic growth, increasing investment in IT infrastructure, significant adoption of emerging technologies, and increasing government initiatives toward the development and deployment of AI technology.



Source: P&S Intelligence, Feb 2020

Top Two Markets for AutoML in 2030



Source: P&S Intelligence, Feb 2020

According to Forrester, every company that works with AI/ML will have a stand-alone AutoML tool in the future. The market is expected to grow substantially as products get better and awareness increases on how these tools fit in the broader data science, ML, and AI landscape. Furthermore, Gartner has also estimated that more than 40% of data science tasks will be automated by 2020.

Currently, the startups in the AutoML market are pushing to increase market awareness and targeting citizen data scientists (anyone who has some kind of an affinity towards the data but is not necessarily a data scientist). There are many AutoML solutions available on the cloud as well, with AWS Sagemaker getting the most attention. With the growth of the cloud, Google Cloud's AutoML and Microsoft Azure's Machine Learning Service are also expected to see plenty of use in the months to come. Moreover, according to Gartner, AutoML is one of the most hyped technologies in AI in 2019.

Challenges in using AutoML

- AutoML systems are still far from being able to solve many of the real-world data science problems where projects are multifaceted and involve large, complex and subjective tasks that do not lend themselves easily to automation. The major problem that AutoML faces is that no single machine learning method performs best on all datasets.
- Another major problem that AutoML faces is Combined Algorithm Selection and Hyperparameter optimization (CASH) problem which aims to identify the combination of algorithm components with the best (cross-) validation performance.
- Lack of transparency into why a model makes a particular recommendation, how the model is selected, etc.
- For now, AutoML is mostly applicable only to Supervised Learning, which has labeled datasets, but not to Unsupervised Learning where there are no labeled datasets.
- AutoML systems are at an early stage of development, where they were first designed to work with the most common data type. Hence AutoML may not successfully work with complex and unstructured data types such as network and web data.
- Even though current AutoML systems are able to handle and process the most common data types, namely, tabular data, text, images and time series, most of the existing solutions are commercial/paid and thus not accessible to everyone.
- Another big challenge will be to see how to make the model search more efficient because people would not like to use a hundred GPUs to solve problems on a small dataset. Companies will need to figure how to make model search process less expensive but without any tradeoff in terms of quality.
- The need for domain knowledge, which is the most prime ingredient in the feature engineering phase, comes at a high price and is which is directly proportional to improved performance. Finding talent with domain knowledge may be a challenge.
- It would be difficult to enable the option to combine automatically generated features with manually created ones despite the need for it.

AutoML Vendor Landscape

There are 3 categories of AutoML solution providers – Open-source, Start-ups, and Tech Giants.

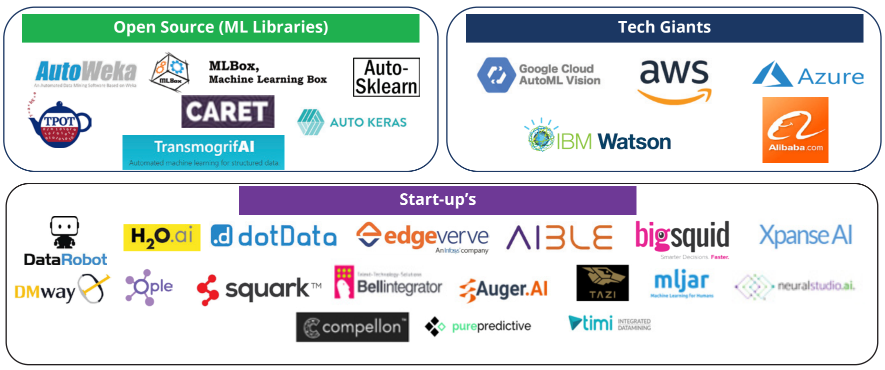


Fig 8: AutoML Market landscape, Feb 2020

Source: Course5 Intelligence

AutoML Market Landscape by Automation Capability

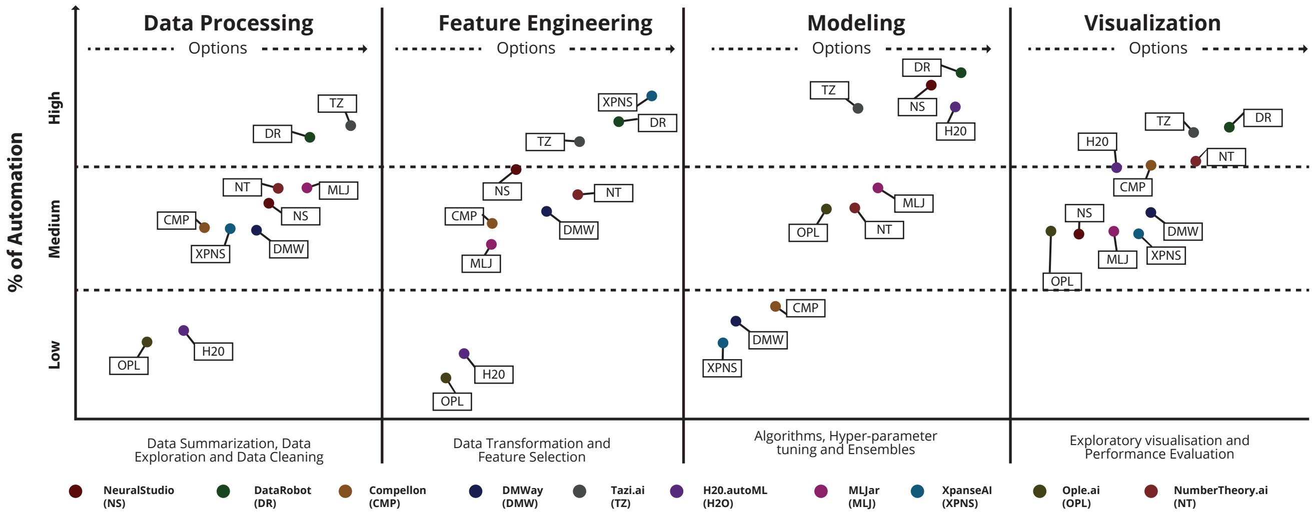


Fig 9: AutoML Market Landscape by Automation Capability

Source: Medium.com

An Overview of Opensource Solutions (AutoML Libraries) in the Market

Automated Hyperparameter Tuning & Architecture Search

- **Auto-WEKA (Free)** – Auto-Weka is focused on helping non-expert users to effectively identify ML algorithms and hyperparameter settings, appropriate to their applications. It simultaneously selects a learning algorithm and enables the setting of its hyperparameters.
- **Auto-sklearn (Free)** – Auto-sklearn performs hyperparameter optimization based on Bayesian optimization. It frees a ML user from algorithm selection and hyper-parameter tuning. Auto-sklearn performs well in medium and small datasets, but it cannot produce modern deep learning systems with the most advanced performance in large datasets.
- **TPOT (Tree-based Pipeline Optimization Tool) (Python) (Free)** – It uses genetic algorithms to optimize ML pipelines. TPOT explores thousands of possible pipelines and finds the one that best fits the data.
- **MlBox (Free)** – It is an AutoML Python library that enables fast reading and distributed data preprocessing/cleaning/formatting. It provides feature selection and predictive modeling.
- **Auto-Keras (Free)** – It is an open source software library that provides functions to automatically search for architecture and hyperparameters of deep learning models.

Automated Feature Engineering/Selection

- **Featuretools** is a python-based library for automated feature engineering. Featuretools works alongside the tools already used by data scientists to build ML pipelines. Users can directly load in dataframes and automatically create meaningful features in a fraction of the time it would take to do so manually.
- **Mlxtend:** Mlxtend is a Python-based library that includes extension and other modules for data analysis and ML libraries.

Tech Giants

Below is a of major tech giants who have AutoML offerings.

- Google Cloud AutoML
- AWS AutoML – Automatic Model Tuning and Amazon AutoGloun
- Microsoft Azure Automated ML
- IBM Watson Studio
- Alibaba

Innovative Start-ups in the AutoML Market

Given below are some of start-ups who already have AutoML offerings catering to some niche areas.

- DataRobot
- H2O.ai
- dotData
- EdgeVerve
- Aible
- Big Squid
- DMWay Analytics
- Squark
- Bell Integrator
- Auger.ai
- Business Insight's TIMi Suite
- Compellon
- MLJAR
- PurePredictive

According to Forrester, DataRobot, H2O.ai, and dotData are the leaders in the AutoML market as they have the best automation capability across all major stages like Data Processing, Feature Engineering, Modelling, and Visualization. EdgeVerve and Aible are strong performers too and exhibit a fair amount of automation capabilities. Big Squid has scope to improve its services, especially the Feature Engineering and Modelling aspects. Bell Integrator, Squark, and DMway Analytics are challengers and need to focus on improving the overall service capabilities. Advanced feature engineering capabilities and model transparency are key differentiators for leaders in the AutoML space, as they empower citizen data scientists to tackle more challenging use cases.

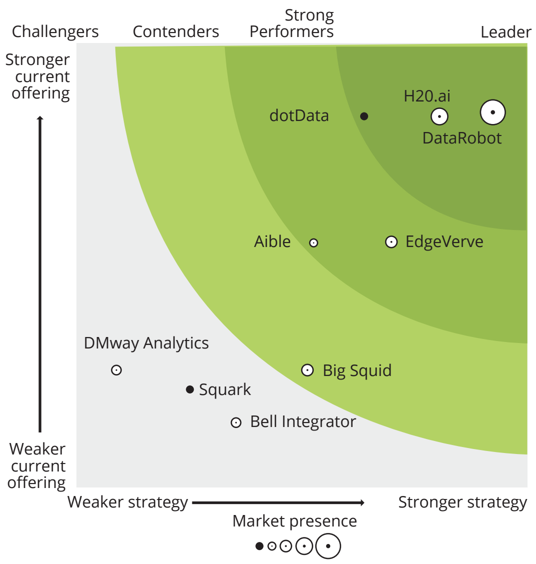


Fig 10: Nine Providers (of AutoML solutions) that Matter Most and How They Stack Up

Source: The Forrester New Wave: Automation-Focused Machine Learning Solutions - Q2 2019

Capability Comparison among AutoML Startups

According to Forrester, AutoML startups mostly have a strong product road map and are focused on in-depth automation of feature engineering tasks. However, these vendors lack the marketing strategy to attract and convert prospects and face lack of visibility among enterprise buyers.

Company	User experience	Data	Feature engineering	Methods	Training	Evaluation	Model operations	Vision	Road map	Market approach
DataRobot	Differentiated	On par	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated
H20.ai	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	On par	Differentiated	Differentiated	On par
dotData	On par	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	Differentiated	On par	On par
Edgeverve	On par	Differentiated	Needs Improvement	On par	Differentiated	On par	On par	On par	Differentiated	On par
Aible	Differentiated	Needs Improvement	On par	On par	Differentiated	Differentiated	Needs Improvement	Differentiated	On par	Needs Improvement
Big Squid	On par	On par	Needs Improvement	Needs Improvement	Needs Improvement	On par	Needs Improvement	Differentiated	On par	Needs Improvement
Bell Integrator	Needs Improvement	Needs Improvement	On par	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	On par	On par	Needs Improvement
Squark	Needs Improvement	Needs Improvement	On par	On par	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	On par	Needs Improvement
DMway Analytics	On par	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Differentiated	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement

Differentiated On par Needs Improvement

Table 3: Vendor QuickCard Overview

Source: The Forrester New Wave: Automation-Focused - Machine Learning Solutions - Q2 2019

Major Initiatives of Big Cloud Players in the AutoML space - AWS, Google Cloud and Microsoft Azure

The current top 3 cloud vendors - AWS, GCP and Azure - started their AutoML journey in 2017-18 and have been building new features in a phased manner. Even though the AutoML services are in the initial stages of testing and cater to niche ML problems, they seem very promising, especially in terms of value addition and cost savings.

AWS

AWS has has a well-defined strategy to target data scientists and developers with its AutoML offerings. The company released Amazon Sagemaker in 2017, a fully managed end-to-end ML service to quickly build, train, and host ML models at scale. Within a year of its launch in 2018, AWS added a new feature to Amazon Sagemaker, which will enable automatic model tuning. Amazon Sagemaker provides one-click training, automatic model tuning and hyperparameter optimization with AutoML capabilities for the most accurate model predictions.

Furthermore, in November 2018, AWS launched Amazon SageMaker Ground Truth, which enables automatic labeling of datasets required for training ML systems. Customers who leveraged this tool reduced labeling costs by up to 70%.

Google Cloud AutoML

Google Cloud has a plethora of AutoML offerings, with a wider capability set than AWS and Microsoft Azure. Instead of training AI/ML models from scratch, Google Cloud AutoML implements automatic deep transfer learning (which means it starts from an existing deep neural network trained on other data) and neural architecture search (which means that it finds the right combination of extra network layers) for language pair translation, natural language classification, and image classification.

Google Cloud AutoML was launched in the Alpha stage in 2017 and followed by multiple sub-products like AutoML Vision, AutoML Video Intelligence, AutoML Natural Language, AutoML Translation, AutoML Tables.

Microsoft Azure's AutoML

Microsoft Azure first introduced its AutoML capabilities in 2018. Azure's AutoML offerings enable designing and running automated ML training experiments and automated model selection, training, and automated hyperparameter settings. Further, Azure's AutoML also simplifies many of the data pre-processing tasks by automatically transforming categorical features into one-hot encoding, imputing missing values and rows, and generating new date-time features.

The Business Case for using AutoML

- Democratizes ML beyond data scientists, making its benefits more widely available
- Increased sales conversions by up to 20%
- Time saved with turnaround in hours or days rather than months, resulting in major cost savings and quick ROI
- Frees Data scientists' time so they can spend less time on feature engineering and hyperparameter tuning, and spend more time experimenting with model architectures
- AutoML pipelines also help avoid potential errors caused by manual work and increase productivity
- Enables non-experts to train high-quality ML models
- Rapid testing and experimenting
- Supports scalable deployment without the need for Development Operationss' deep ML knowledge
- Improves the efficiency of finding optimal solutions to machine learning problems

Some Areas that cannot be automated by AutoML

- Real-world problem identification
- Formulating the problem as a data science problem – whether the problem needs to be addressed as a Supervised, Unsupervised or Reinforcement Learning task, or as traditional statistics
- Risk anticipation and solutions to manage them

- Data collection design and assessing the data quality, especially unlabeled data
- Managing and controlling human biases
- Incorporating domain knowledge into the process
- Analyzing ethical issues and assessing the impact of project output in society
- Effectively communicating results to stakeholders

Future of AutoML

AutoML systems are headed towards becoming mainstream in the Machine Learning world. Gartner predicts that over 40% of data science jobs will be automated by 2020. This does not mean that AutoML will replace Data Scientists in the future. Instead, AutoML will be leveraged as an assistant tool, which will lower dependence on the data scientists, by iteratively trying out an algorithm, scoring its performance, and choosing and refining other models. Eventually, AutoML solutions will evolve to platforms that automate more of the data science workflow and may be offered as AutoML-as-a-solution. With the deployment of AutoML, data science roles will become more of business science roles. Plug-and-play AutoML solutions will become a must-have for every organization looking to scale up their usage of ML.

According to Forrester, most organizations can benefit from a standalone AutoML solution, and it is expected that this market will grow substantially as products get better and awareness increases regarding how these tools fit in the broader data science, ML and AI landscape.

Developments in the AutoML space are more mostly focused on improving the speed of generating production deploymentready models, rather than exploring how the technology can be improved for more difficult problems like fraud and network intrusion detection, etc.

In the future, the driver's seat in AutoML products will be occupied by business users like Analysts and Consultants, where users can add their knowledge to aid the machine in generating very accurate models automatically. Technologically, the growth of AutoML will be in the hands of software developers; however, accuracy, which is one of the most important aspects of Machine Learning, will be aided by business users. Business users will continuously be involved in fine-tuning ML solutions with their business knowledge. Some of the complex data science pipelines will become increasingly streamlined, and a large variety of algorithms will be added to aid better optimization.

Below are the few areas where AutoML can provide better solutions in the future:

- **Data Cleaning Management:** Existing AutoML tools in the market do not handle the end-to-end data cleaning process. AutoML aims to automate some part of the unstructured data cleaning process and make it ready for analysis.
- **Improved Deep learning:** AutoML is most often referred as Deep Learning method, which abstracts all the complex parts without having to design complex deep networks. With just the data, AutoML can easily find architecture quickly, automating DNN architecture design and the preprocessing of data prior to modeling.

- **Ability to scale large datasets (Big data):** The existing AutoML tools in the market have not been designed for Big Data platforms. They usually are based on Python or Java, so one could use them with the Python or Java-bindings of those platforms, but they are a little slow while handling Big Data. Hence Spark, TensorFlow, etc., are expected to bring scalability to AutoML. AutoML and Big Data platforms can benefit from tighter integration in the future.

- **Manage unlabeled data:** There is plenty of potential in Unsupervised Learning which can unlock loads of new opportunities.

- **Interactive platform:** In the future, AutoML may become more of an interactive platform where the user and AutoML system will work responsively together.

Business Use Cases of AutoML

- **Feedzai**, which provides software to fight fraud with AI in the banking and commerce sector, is leveraging AutoML to automate tasks like feature engineering and machine learning model creation. Data scientists are now able to create fraud prevention solutions as much as 50 times faster than what is possible with the traditional data science workflow.

- **Facebook** trains and evaluates a vast number of ML units every month (roughly 300,000 currently). The company essentially built an ML assembly line to take care of all its models. Facebook has even built an AutoML engineer (called Asimo) that automatically produces improved versions of current versions.

- **Mercari** is a popular shopping app in Japan that has been using AutoML Vision (Google's AutoML solution) for classifying images with an accuracy rate of 92%.

- **Disney** is using Google's Cloud AutoML to build vision models to annotate its products with Disney characters, product categories, and colors. These annotations are being integrated into Disney's search engine to enhance the impact on Guest experience through more relevant search results, expedited discovery, and product recommendations on shopDisney.

- **Urban Outfitters** is experimenting with Google Cloud AutoML to automate the product attribution process by recognizing nuanced product characteristics like patterns and neckline styles. Cloud AutoML holds great promise to help Urban Outfitters' customers with better discovery, recommendations, and search experiences.

- **Chevron Corp** is an early adopter of Google's AutoML technology. Chevron's seismic imaging and processing team implemented the alpha version of AutoML Vision image analysis tool to see through internal documents to evolve new opportunities for oil drilling.

- **The Department of Homeland Security (DHS) Science and Technology Directorate (S&T) in the US** announced that DataRobot, Inc. received $200,000 to begin testing a prototype of an Automated ML platform for the U.S. Customs and Border Protection's (CBP) Global Travel Assessment System (GTAS) in 2018. DataRobot proposes to apply automated machine learning (AML) to GTAS to expedite the model development process.

- **Huawei** has developed the industry's first AutoML platform for mobile communications which has the ability to reduce the data modeling cycle from months to days, and also accelerates the incubation of Wireless AI applications.

How AutoML will Impact Business Outcomes

Businesses have just begun to experiment with AutoML and are beginning to apply it across some of their minor tasks and assess the accuracy levels of the output of AutoML systems. Currently, there is no universally best AutoML approach.

AutoML has the ability to empower Product Managers to transform their thinking towards real-world business problems. Once a problem is solved with AutoML, the knowledge gained while solving such problems is recorded and in future can be leveraged to resolve a different but related problem, in less or no time.

For example, the knowledge gained while learning to recognize cars could be applied when trying to recognize trucks. AutoML is gearing to solve many such business problems across various business horizontals like Sales and Marketing, Research and Development, Supply chain, Logistics, Production, Cybersecurity, etc.

Earlier, each Product Manager across different verticals had specific tasks to analyze data in their respective verticals and come up with solutions, methods and future recommendations to solve specific problems or develop new functionalities. In order to address this, Product Managers leveraged Data Science and Machine Learning to come up with a wide range of solutions. The Product Managers addressed these issues in isolation with their own unique methodologies and came up with different solutions, insights, and recommendations for their specific vertical.

In order to analyze data, every time the data has to be prepared, labeled and fed into the ML system. Then the feature engineering and model/algorithm selection is performed, followed by testing and final output. However, traditionally the processes ended there, and this process, along with the output, was not recorded. Whenever such events occurred, the Product Managers had to again run all the processes to develop new insights. These yearly or quarterly repetitive exercises used plenty of time and resources, resulting in slow response. Furthermore, the data, methodologies, output, insights, tested practices and the entire learning process rest only with the specific product manager, which was never reused nor shared amongst other verticals. This created a gap in cross-learning amongst different verticals and failure to reuse data.

AutoML aims to close this gap and aid in cross-learning amongst different verticals within enterprises. AutoML can learn from one process and implement it in other processes that are similar in nature. This will not only help to solve specific business problems across different teams faster but will also keep the teams aligned to one single goal.

Conclusion

In the current era, the rapid growth of data is outpacing the ability to analyze and make sense out of it. Currently, Machine Learning seems to solve some of these challenges with data scientists playing a major role. In the next 2-3 years, with the advent of emerging technologies like 5G, IoT, Smart cities, etc. there would mammoth amounts of rich data at one's disposal, which would require a combination of powerful tools and human touch in a 70:30 ratio to make sense of the data. This emerging scenario definitely needs an automated approach and AutoML seems to be the perfect solution.

The AutoML market is still in its early stages right now but has the potential to grow very rapidly, especially with a major push from large companies like Google, Amazon, Microsoft, and Alibaba. Several startups have also been adding to the industry knowledge base, especially in niche business areas.

As the industry is moving towards finding the best AutoML approaches, it requires practitioners and data scientists to provide feedback on the right tools to for a robust AutoML ecosystem.

About Course5 Intelligence

Course5 Intelligence enables organizations to make the most effective strategic and tactical moves relating to their customers, markets, and competition at the rapid pace that the digital business world demands. We do this by driving digital transformation through analytics, insights, and Artificial Intelligence (AI). Our clients experience higher top line and bottom line results with improved customer satisfaction and business agility. As we solve today's problems for our clients, we also enable them to reshape their businesses to meet and actualize the future.

Rapid advances in Artificial Intelligence and Machine Learning technology have enabled us to create disruptive technologies and accelerators under our Course5 Intelligence suites that combine analytics, digital, and research solutions to provide significant and long-term value to our clients.

Course5 Intelligence creates value for businesses through synthesis of a variety of data and information sources in a 360-degree approach, solution toolkits and frameworks for specific business questions, deep industry and domain expertise, Digital Suite and Research AI to accelerate solutions, application of state-of-the-art AI and next-generation technologies for cognitive automation and enhanced knowledge discovery, and a focus on actionable insight.

Visit : www.course5i.com

Course5
Transformative intelligence