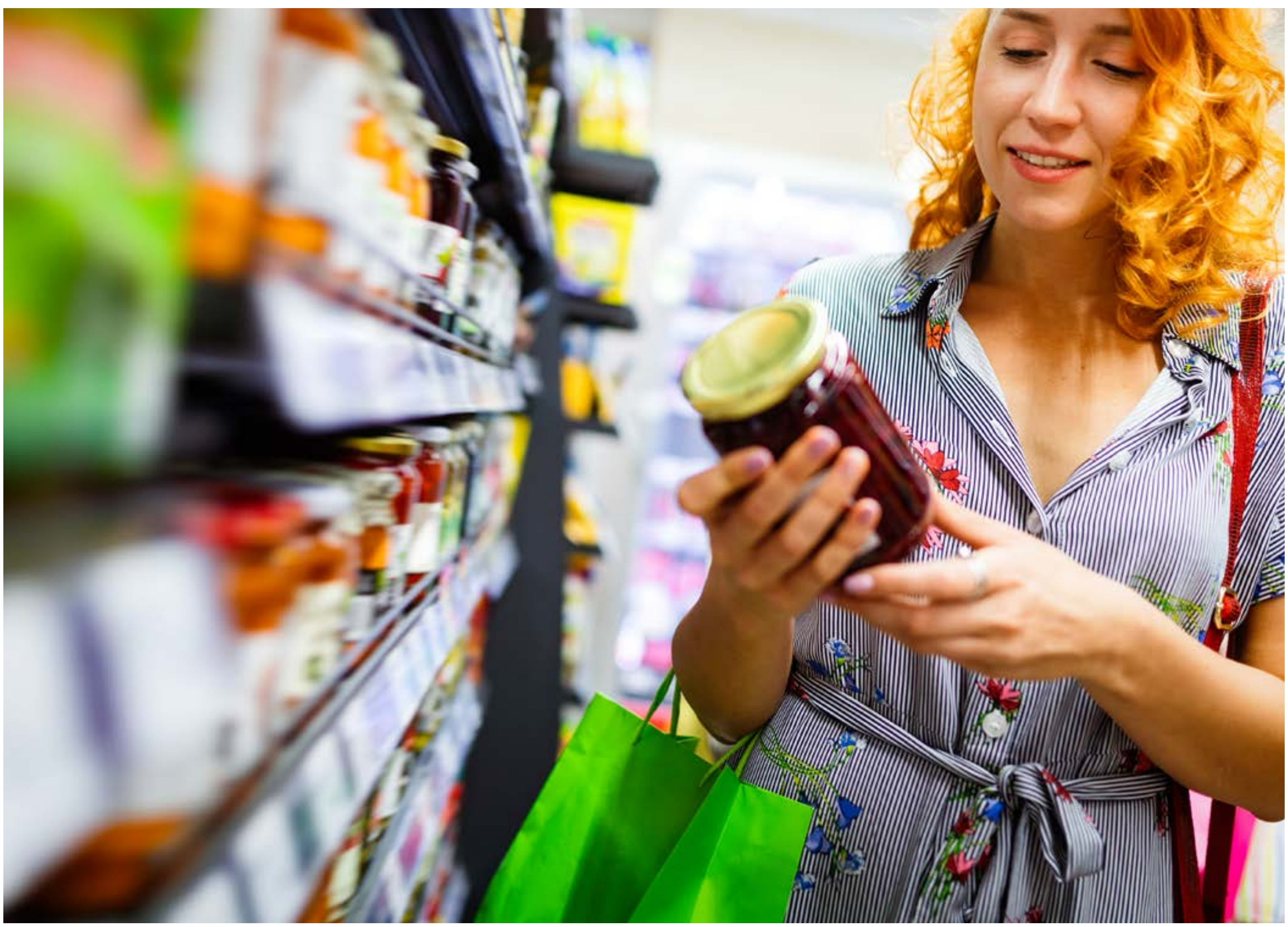
MACHINE LEARNING FOR RETAIL MANAGEMENT

Creating a Data Driven Business



$.9T

$.9 – $1.7 trillion value being added to Retail annually by AI/ML

An Opportunity That Can't Be Ignored

The economic impact of machine learning (AI/ML) for Retail and Consumer Packaged Goods use cases is impressive—with an estimated $.9 - 1.7 trillion value being added to Retail annually. In fact the retail segment has the highest potential to benefit from ML use cases when compared to all individual industry segments. This immense value is being played out in all segments and in all points of the value chain, with the potential upside for advanced analytics in this sector set to redefine Retail as we know it.

However, deploying and scaling models that power AI/ML use cases across the enterprise requires implementing complex, iterative workflows from data to models to business outcomes. This is not easy. In fact, currently only 35% of organizations indicate that analytical models are fully deployed in production.[1] In addition, as the number of AI/ML projects and models in production multiply, maintaining your ML use cases at scale can be slow, cumbersome and fraught with "false starts"—not to mention the substantial operating overhead often needed to keep models accurate in the long-term. For many retail businesses, implementing ineffective tools and approaches to ML can mean expensive mistakes that can seriously limit your return on investment.

As a line of business leader, the upside potential can't be ignored. By understanding the basic principles and use cases for machine learning and the resultant data driven business insights — profitability, agility and distinct competitive advantage — will be delivered.

Coming to Know Machine Learning

McKinsey defines machine learning as, "...algorithms that detect patterns and learn how to make **predictions and recommendations** by processing **data** and experiences, rather than by receiving explicit programming instruction. The algorithms also adapt in response to **new data and experiences** to improve efficacy over time."

Reflecting on three points from the definition will help uncover meaningful foundational principles. **Predictions and recommendations** are the core foundational value of ML, as it gives your business the ability to make predictive and prescriptive decisions both in the short and long term. This provides recommended actions that are beyond what conventional analytics or human analysis can perform. **Data** is the foundation of machine learning and the more varied, accessible, unified, integrated, streamlined and secure your data pipelines are, the better your ML models can predict and prescribe business decisions that lead to positive outcomes. It has to be understood too that ML models should never be static or unattended, but need monitoring, updating and inspection **when new data and experiences** are seen, thus keeping them accurate and relevant. Static models will skew, drift and deliver degraded—and often flat-out incorrect—business recommendations.

COMMON RETAIL TRANSFORMATIONAL USE CASES THAT LEVERAGE AI/ML

PERSONALIZED INTERACTIONS

Customer profiles	Customer engagement, Conversion analysis
Recommendations, Next best offers	Relevant, contextual content, messages
Social listening, engagement	Intelligent call center response

CUSTOMER-CENTRIC MERCHANDISING

Dynamic pricing	Distinctive product Development
Localized assortment	Promotional planning & forecasting
Demand forecast accuracy	CPG partner collaboration

SUPPLY CHAIN AGILITY

Inventory	Distinctive product Development
Localized assortment	Promotional planning & forecasting
Demand forecast accuracy	Equipment predictive Maintenance

REIMAGINING STORES

DYNAMIC PRICING	Distinctive product Development
Localized assortment	Promotional planning & forecasting
Demand forecast accuracy	Equipment predictive Maintenance

Machine learning is being applied to all points of Retail's value chain. Some of the direct benefits that ML provides are:

- Enabling growth and expansion of product lines at scale due to a more optimized processes and supply chains
- Capital cost reductions through predictive maintenance
- Improved supply chain management through effective inventory management and well monitored and synchronized product flow
- Improved human-robot collaboration improving employee safety conditions and boosting overall efficiency
- Loss prevention, shrink through streaming analytics that identify and alert in-store associates of potential (real-time) fraud and shrink
- Dynamic pricing providing the ability to change prices considering both competitive pricing and predictive customer response models
- Responsive customer-focused merchandising offering quick changes to predictive market demand
- Providing the next best offer, optimizing pricing and promotions through personalized marketing
- Increase conversion and customer engagement with deeper personalization and targeting through personalized interactions
- Synchronized Customer Service by providing stores and call centers with 3600 view of customer interaction across all touchpoints and recommend next best action
- Developing smart products - A smart, or connected, product is a device that is linked to the Internet so it can share information about itself, it's environment and its users and make predictions that modify performance

Data is Delivering Value to the Business[2]

The upside potential of deploying AI/ML in your Retail business is substantial and an initial step is embracing and understanding which use case will return the most value. Retail and Consumer Packaged Goods naturally place different emphasis and priority on various use cases, as Retail is focused on driving value through ML through its four strategic pillars: personalized interactions, customer centric merchandising, supply chain agility, and reimagining stores. Consumer Packaged Goods have added dimensions with product development and operations.

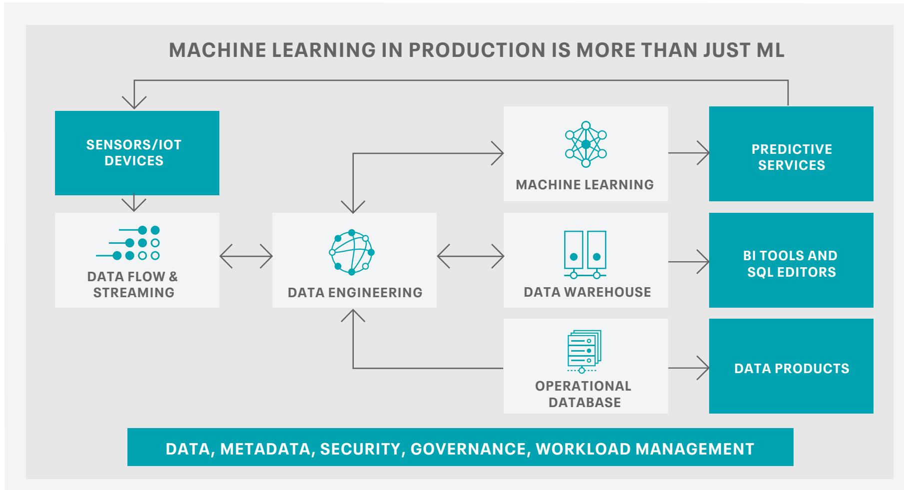
SEGMENT	TOTAL DOLLAR IMPACT	POTENTIAL IMPACT as a percentage of Industry revenue	ANNUAL BUSINESS IMPACT BY USE CASE		
Retail	$900B-$1,700B	8.7%	1 Pricing and Promotion $354 - $700B	2 Inventory Optimization $121 - $238B	3 Customer Acquisition and Lead Generation $96 - $186B
			4 Task Automation $79-$158B	5 Customer Service Management $$82 - $121B	6 Logistics Network and Warehouse Optimization $73B
Consumer Packaged Goods	$700-$1,400B	9.4%	1 Yield, Energy & Throughput $170 - $387B	2 Inventory Optimization $123 - $238B	3 Marketing Budget Allocation $101 - $202
			4 Predictive Maintenance $89-$180B	5 Channel $30-$50B	6 Pricing and Promotion $23- $35B

Data Drives AI and Machine Learning

The following table demonstrates several hypothetical representative machine learning use cases in the four strategic pillars of Retail:

USE CASE EXAMPLE	DATA INPUT SOURCES	RESULTANT VALUE TO THE BUSINESS
Personalized Interactions - Intelligent Call Center Response	• Speech-to-Text • Call Center Logs • Social Media • Offer History • Conversion • Fulfillment Preferences • User Click Streams	Higher customer satisfaction resulting in next best offer conversion
Customer Centric Merchandising - Demand Forecast Accuracy	• Geo-Political/New Events • Search Marketing (SEO) • External Demographics • Promo (Lift) Effectiveness • Weather /Local Event Calendars	Less stock outages combined with resulting higher conversion
Supply Chain Agility - Predictive Asset maintenance	• IoT Sensors from Vehicle / Equipment • Geo-Location • Satellite/Traffic • Supply Chain Scheduling • Demand Forecast • Streaming Video/Computer Vision	Capital asset downtime minimized by optimally scheduled and lower cost, maintenance
Reimagining stores	• Streaming Video/Computer Vision • Proximity Beacons • RFID/Sensors • Social Media • POS Transactions • Demand Forecasts • Fulfillment Preference	Higher customer satisfaction with BOPUS, higher conversion due to lower stock outs, and lower supply chain and fulfillment costs with optimize supply chain

The table illustrates that data comes from the entire enterprise; instore sources such as point of sale, beacons, RFID tags, streaming video and computer vision; external sources such as demographics, social media, satellite and traffic and geolocation; order history or product preferences such fulfillment preferences, offer details, POS transaction detail, and user click stream. Data comes in several forms - dimensional or structured data, that can be typically organized in data tables and arrays; semi-structured data such as user click stream, omni-channel pricing, demand forecast; and also behavioral or unstructured data such as demand forecast and edge data, streaming video/computer vision and call center captured voice that usually is streaming and has a time dependency to it.



Data too could reside in multiple locations - on prem, hybrid and cloud; could be shared by other advanced analytical tools running your business such as Operational Databases (OpDB) that control and run essential business operations in real time, or data warehouses that plan, solve problems, support decisions, discover hidden insights; and lastly data utilized for machine learning that can be both streaming and batch oriented.

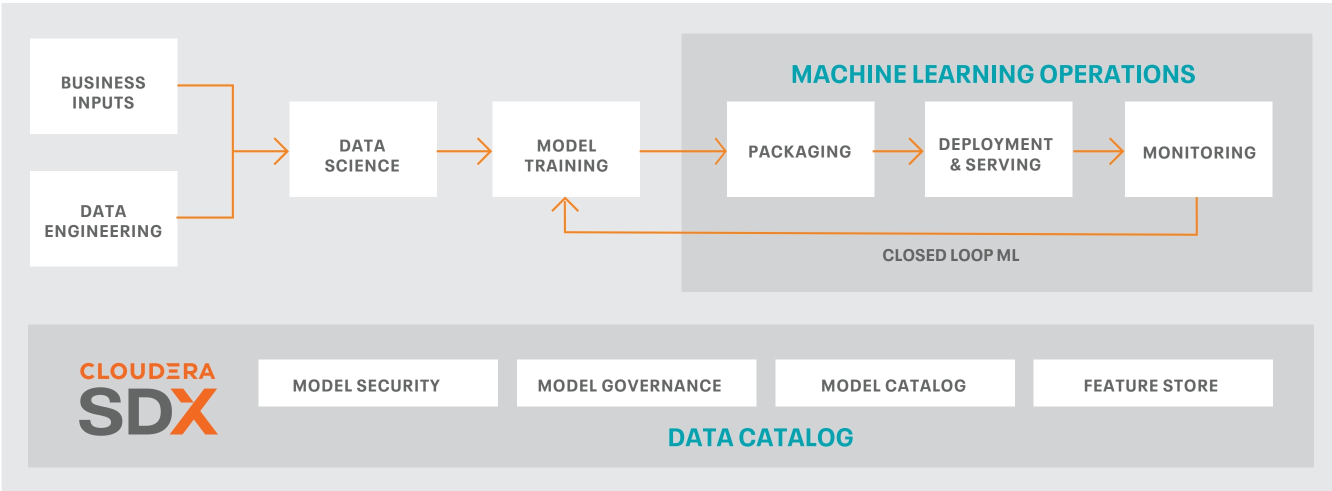
Challenges That Accompany Data

Some of the challenges that accompany data when building ML models are:

- **Inability to converge structured, semi-structured and unstructured data:** As was illustrated earlier, data comes in all forms and for an enterprise to execute advanced analytics and machine learning, the organization needs to leverage all sources in all formats
- **Siloed data sources:** Your business most likely runs a mix of data warehouse and storage locations (on-prem, hybrid or cloud). Applying data for advanced ML analytics requires the ability to access data from any location. Most businesses use a blended technology solution to mitigate risk and lower costs necessitating the need for flexible access capabilities.
- **Point solutions:** Machine learning applications are not stand-alone solutions in your business, but work with and complement other business applications across the full data lifecycle to effectively utilize your data sources and streams. Stand-alone applications require integration with many other products to fill gaps such as data ingest, data engineering, production ML, and business applications. Coupled with random release schedules and managing multiple systems and data silos with the propensity to break the chain of IT security and governance are created.
- **Inability to ingest data-in-motion and batch oriented data:** Advanced analytics require data in any form from any source in both streaming and batch format. Limitations of data ingestion, analysis, or storage hinders business insight by limiting the potential of insight of machine learning by restricting the use of enterprise data when building the ML model.

A Deeper Dive into Machine Learning

Creating and bringing a machine learning model into production from an on-prem, hybrid or cloud architecture can be defined by the following steps:



Data Engineering consists of data acquisition, processing data pipeline automation, and governance. Streaming data contains some portion of the data that is not suitable for use. Consider a produce freezer or cold storage start-up or abrupt shut down - this data could contain spikes in sensor readings that are atypical of the process, hence they need to be removed. Data processing cleanses and prepares raw data for a developer to run analysis, as is typically not in a convenient format. As organizations use data (and analytics) more often, and for more important questions, the need to govern those assets increases. Every organization should be concerned about data quality in their source systems, but often these concerns are isolated and not visible across departments. Security, privacy and regulatory compliance are important elements of governance.

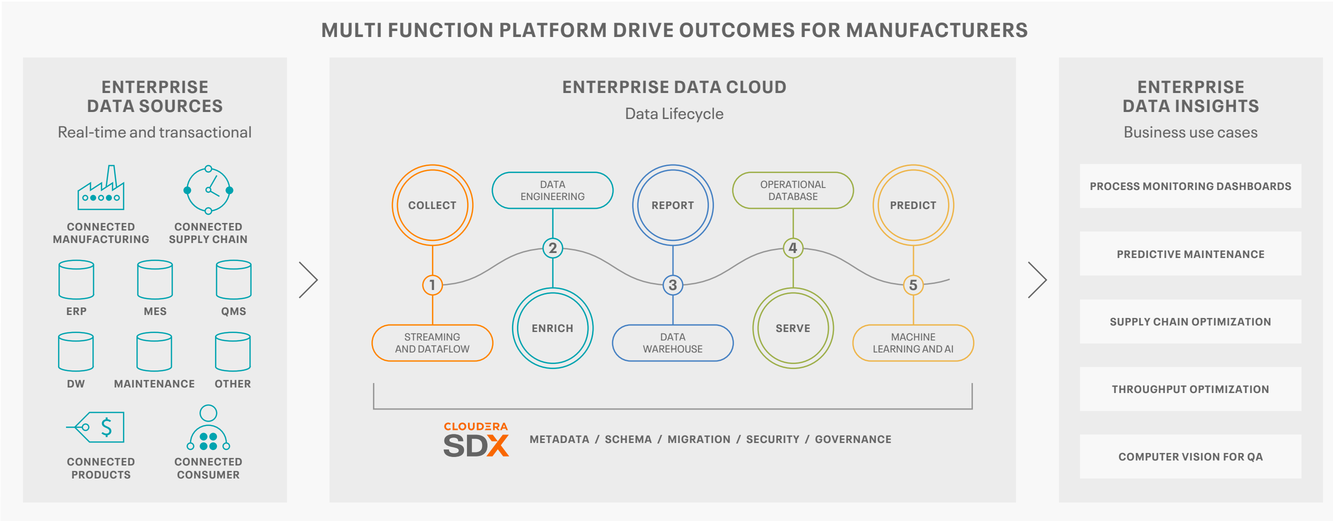
Data Science includes exploratory analytics, experimentation, and machine learning model building for use in production environments and in predictive applications. Exploratory analytics enable data scientists to deeply understand their data and the potential it has for advanced analysis and machine learning. Experimentation is a component of discovering what problems the data can solve, and this work enables the building and testing of predictive models. These models are then delivered into the business either through data applications, automated processes, or business data visualizations and dashboards that enable business teams to better make strategic decisions.

Production ML is the process of delivering, monitoring, and sustaining a working advanced analytic model (better automation, predictions, innovations, etc.) to stakeholders (customers, internal business, etc.). There are several different ways to deploy a model and understanding the end user (customer) intention helps determine the technology required. The deployment phase can be as simple as generating a report or as complex as implementing a repeatable data science process that automates an operations step, allies computer vision to quality control or recommends a “next best offer” to a final customer.

Monitoring. As hard as we try to prevent them, production processes change over time - sometimes for the good or more often in a way that we can’t explain. Select sensors or beacons in a store could be relocated resulting in subtle changes - customer flow in a store now has gaps or a social platform has changed concerning user demographics (Boomers using Facebook vs Gen Z with Tik Tok). Monitoring is an essential element of production deployment as it provides visibility into the accuracy and viability of the machine learning model deployed. Poor visibility into mathematical metrics and to the external tools used for monitoring is a major challenge. This challenge can be mitigated by using a unified model monitoring platform for all the deployed models from a single pane of glass and having a clear understanding of the data sources, mathematical models used to build the ML model or how other models are performing in the enterprise. Additionally, it's important to select a platform that enables automatic ground truthing for accuracy and drift, which in turn power automated model maintenance workflows such as retraining—enabling greater scale with less operating costs overall.

SUCCESSFUL RETAIL USE CASES EMBODYING MACHINE LEARNING

	Problem	Solution	Business impact
DEMAND FORECASTING, AUTOMATIC REPLENISHMENT **A German international discount supermarket chain that operates over 12,000 stores across Europe and the United States**	**The inability to respond accurately and just in time to changes in demand and supply:** • sales forecasts based on intuition and manual analysis of fragmented data • manual generation and alteration of orders and store replenishment	**AI-driven analytic platform created and implemented by solutions provider:** • Advanced daily sales forecasts at an item level, based on daily receipt sales, supported by both internal attributes (e.g. marketing campaigns) and external attributes (e.g. weather, seasonal sales) • **32M forecasts per day automatically generated replenishment commands** and orders for supplier based on advanced sales forecasts (up to 90% of standard products delivered automatically without human, manual intervention)	**Elimination of manual and intuitive ordering of products to stores resulted in:** • **3% to 7% decrease in losses** by reducing overstocks • **30% to 50% reduction** of average out of stocks in stores
REAL-TIME FULFILLMENT, ROUTE OPTIMIZATION **A British multinational groceries and general merchandise retailer that is the ninth-largest in the world measured by revenues**	**Legacy systems were falling short of delivering on omnichannel, predictive analytics and enterprise strategic needs:** • Existing route optimization tools not reflective of 'true' drive time for deliveries • Delivery van fleet maintenance was scheduled, resulting in 'unplanned downtime' and the need for 'back-up' equipment to ensure customer experience standards were always met 'at all costs'	**Cloudera CDP, a scalable, enterprise data science/analytic platform was used to integrate and collect all operational data into HQ data lake. Results are as follows:** • Improved data access on omnichannel orders, logistics, delivery capacity • Brought IP in-house, reducing point solution costs, improving analytic agility and automation of production analytic into existing business applications • Implementation of predictive analytics for maintenance (v. scheduled)	**The enterprise data lake resulted in:** • Improved intra-day online order fulfilment demand **forecast accuracy by 3%** • Improved customer delivery capacity, service with shortened delivery windows • Reduced number of vehicles / drivers by 140 (@ $150k cost per) = **$21m**
REAL-TIME SHRINK (LOSS) PREVENTION **A world leading American multinational retail corporation that operates a global chain of hypermarkets, discount department stores, and grocery stores.**	**There was a need to improve cycle-time to action on cold storage temperature fluctuations. Objectives were to:** Create an enterprise-class IoT/big data solution • Enhance notification and response times on potential shrink candidates due to temperature fluctuations impacting fresh product life-cycles • Enable NRT predictive maintenance analytic capabilities for cold storage (chilled and frozen) to remediate occurrences of equipment failure leading to losses	**The solution was an intraday cold storage sensor data collection, analysis and store 'action' messaging infrastructure. It provided:** Ability to ingest data from 1000's of cold storage sensors across all stores into data lake HDP (no requirement for structure/schema definition prior to import) • Streaming data ingestion of temp. readings, OH inventory and product lifecycle benchmarks for highly perishable food products (e.g. frozen food, produce, meat) • Rapid query response times using established business rules, predictive analytics • NRT store communications, messaging on product and potential equipment risks	**Projected business impact:** Combined **$250M** in equipment maintenance costs and product shrink reduction annually



A Successful Data Driven Manufacturing Business

Machine learning can't be seen as a separate process or entity in your data strategy, so don't consider machine learning in a vacuum. When designing your data strategy consider the following:

Infrastructure Technology

- The need for real time open data access - modern retail organizations have complex infrastructure strategies (on-prem, hybrid and cloud) and data is access is needed from all sources
- Point solutions require an increased dependence and effort ensuring connectivity. The persistent threat of breaking IT security and governance is also a strong possibility

Use Case Sustainability

- The need for monitoring to provide full visibility of model skew, drift and accuracy changes ensures model accuracy and insight
- Data is shared with other business tools, so understanding your machine learning catalog - what models are built and what they are doing is essential
- The fact that taking the model from concept to production is sometimes hard. The compounding factor is scaling a handful of production models to tens or hundreds

Business Sustainability

- Creating and deploying machine learning can be an internal activity, or can rely upon Independant Solution Providers or software vendors. Cloudera recognizes the dynamic environment of the fast pace changes in technology. Cloudera addresses this environment with Cloudera Fast Forward Labs Research focusing on emerging trends that are still changing due to algorithmic breakthrough, hardware breakthrough, technological commoditization, and data availability. Cloudera supports your IT team with reports, technology perspectives and working prototypes that exhibit the capabilities of the algorithm and offer detailed technical advice on its practical application, thus augmenting your technical abilities

Why Cloudera

Cloudera Data Platform enables financial services providers to effectively execute their data and analytics strategy to address current and evolving customer expectations.

EDGE TO AI ANALYTICS
All the functions needed to ingest, transform, query, optimize, and make predictions from data are integrated, eliminating the need for costly point products.

DATA SECURITY & COMPLIANCE
Maintains strict enterprise data security, governance, and control across all environments.

HYBRID AND MULTI-CLOUD
Delivers the same data management capabilities across data centers, private, and public clouds.

100% OPEN SOURCE
Open compute and open storage ensures zero vendor lock-in and maximum interoperability.

About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. We deliver the modern platform for machine learning and analytics optimized for the cloud. The world's largest enterprises trust Cloudera to help solve their most challenging business problems.

Learn more at cloudera.com

Delivering Value Through Data

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. Whether your organization wants to build advanced analytics machine learning use cases that personalize interactions, improve customer centric merchandising, embrace predictive maintenance, add agility to your supply chain or reimagine your stores - we are here to help.

The Cloudera Data Platform (CDP) is an enterprise data cloud for any data, from any source, that runs in any cloud, on-premises or hybrid environment. CDP and the Cloudera Machine Learning (CML) experience brings data that runs the business and data generated from the manufacturing process into one platform where enterprise-wide machine learning use cases can be leveraged for real time business insights and decisions. CDP and CML enables the business with:

- An end-to-end data management and analytics platform that can help manufacturers drive insights and action from any data, anywhere, in real-time.
- The ability to ingest, process and analyze high volumes of real-time data from any source—connected equipment, production sensors, computer vision, historians, ERP & MES systems, historical archives, master data management databases, fleet vehicles, or worker wearables.
- Offer massively distributed storage and processing engines for large data sets to execute a wide range of data processing workloads.
- Enable predictive analytics or apply machine learning algorithms to petabytes of data, while maintaining strict enterprise data security, governance, and compliance, audit trails across on-premise and cloud hybrid environments.
- Glean insights from unstructured data sources originating from process sensors, computer vision, robotics, or acoustic sensors.
- The ability to build, test, iterate, and deploy machine learning models to enable use cases such as predictive asset maintenance, demand forecasting, or intelligent call center response.
- Provide multiple analytical options to drive insights, intelligence, and action from data at the edge, on premise, or in any public, private, or hybrid cloud.

Cloudera is deeply involved in the transformation of the Retail Industry. The case studies and examples are just a sample of the work we are doing in marketing and sales, supply chain optimization and CPG product operations with the Cloudera Machine Learning experience. To test drive Cloudera Machine Learning or learn more, just click here.

1) IDC's Advanced and Predictive Analytics survey and interviews, n = 400, 2017 – 2019.

2) https://www.mckinsey.com/featured-insights/artificial-intelligence/visualizing-the-uses-and-potential-impact-of-ai-and-other-analytics

CLOUDERA

Cloudera, Inc. 5470 Great America Pkwy., Santa Clara, CA 95054 USA cloudera.com