ARTEFACT

DATA & AI
FOR RETAIL

ARTEFACT

AI IS ABOUT PEOPLE

WE ACCELERATE DATA AND AI ADOPTION TO POSITIVELY IMPACT PEOPLE AND ORGANIZATIONS.

The Netherlands
UK
Germany
France
Switzerland
Spain
New York
Lebanon
South Korea
Morocco
China
Senegal
Saudi Arabia
Los Angeles
Mexico
UAE
Colombia
Malaysia
Singapore
Brazil
South Africa

20
COUNTRIES

1500
EMPLOYEES

+1000
CLIENTS

Artefact is a global leader in consulting services, specialized in data transformation and data & digital marketing, from strategy to the deployment of AI solutions. We are offering a unique combination of innovation (Art) and data science (Fact).

DATA & AI ACCELERATION PROGRAMS | AI SOLUTIONS | DATA MARKETING

TABLE OF CONTENTS

Data & AI for Retail

The outlook for data and AI transformation, today and tomorrow.

An interview with Vincent Luciani

The generative AI technology revolution has been a paradigm shift for all industries and sectors. Artefact sees AI as an incredible opportunity that, if used properly and ethically, will lead to economic, social, and democratic progress.

"At Artefact, we hold an optimistic vision, viewing AI as an incredible chance that, if used properly and ethically, will lead to economic, social, and democratic progress."

How is generative AI profoundly transforming society and businesses?

We are at the beginning of a new era. The generative AI revolution is reshaping societal and economic landscapes. After an experimentation phase, generative AI will continue to change the game for the global community. It's a technology with the potential to improve the world in many ways, as long as solid checks and balances are in place to ensure its responsible and beneficial development.

- **Economically,** it offers undeniable productivity gains that will spur innovation and new business growth.

- **Socially,** generative AI will streamline administrative tasks, freeing up more valuable and creative time, which could lead to innovative job opportunities and the development of new skills.

- **Democratically,** the accessibility of GenAI to all will provide deep knowledge and solutions to address specific societal and educational inequalities and advance the cause of social justice.

How is Artefact leading the generative AI transformation for enterprises?

Since the availability of the first LLM models (Large Language Models), even before the official public launch of ChatGPT in November 2022, we at Artefact have been one of the key global pioneers using this powerful technology, designing and deploying many generative AI use cases with our clients throughout 2023.

As certified experts with major Clouds and open source GenAI, we've already acquired strong expertise and developed a solid ecosystem. In this context, we recently announced our official strategic collaboration with Mistral AI, the most powerful LLM platform for a European OpenAI.

Despite achieving notable reductions in development time and enhanced employee adoption, scalability of GenAI projects remains a challenge, emphasizing **the need for ethical and secure environments grounded in robust data foundations.**

For more than 10 years, Artefact has prioritized **the crucial role of data in AI success for enterprises.** Initiating data acceleration programs, we focus on elevated data quality, governance, and interconnected platforms, adhering to ethical and responsible guidelines.

Anticipating substantial growth thanks to these new LLM technologies, companies are urged to **embrace AI for a competitive edge.** This transformative year will necessitate new organizational models and widespread AI deployment across business value chains, with Artefact accompanying its clients from strategy to full operations.

However, the success of technology shifts depends on **fostering trust and enthusiasm among all employees,** requiring consultation and support from top to bottom, an area where hackathons and training can be instrumental.

What is Artefact's mission? What initiatives have you taken to anchor your AI adoption strategy to accelerate business growth and efficiency?

Artefact's primary mission is to accelerate the adoption of data and AI to positively impact people and organizations.

To translate this purpose, our motto is **«AI is about people».**

The companies that will endure are those that successfully foster **a data culture** with access to knowledge and data for all.

We've undertaken several initiatives in this area that are highly strategic to Artefact's positioning as **a major player in data democratization,** in order to fully realize its potential for positive transformation.

- The development of the **Artefact School of Data,** a key pillar in our strategy of providing clients with training adapted to the constantly evolving skills of the data industry. We are also developing «à la carte» e-learning platforms for clients to quickly share knowledge of data and AI with all of their employees. We've expanded our **Artefact School of Data** from France to Dubai and New York City, and soon to other cities to educate organizations about data and AI, while creating new job opportunities in this domain.

- We've also launched many **generative AI hackathons** at major companies to empower and inspire their employees with these new innovative technologies.

- For over five years, we've organized **large-scale conferences** such as **AI for Finance & Industry** and **AI for Health.** We also successfully launched the first edition of **AI for Luxury in NYC** and **AI for Life in Geneva,** bringing together top-level AI ecosystem stakeholders, including

With significant growth expected thanks to generative AI technologies, we advise all organizations to embrace AI to gain a competitive advantage.

major corporations, startups, and universities, to disseminate knowledge about data and AI.

Business data maturity has advanced rapidly over the past decade. How has Artefact evolved as a global leader in data and AI consulting services?

Companies have implemented **data governance policies,** which are a prerequisite for any transformation, but there are still sectors that lag far behind in terms of their data processing, with a real potential for efficiency.

We started to transform marketing departments by making them more profitable and relevant in their multi-channel media investments with pioneering targeting, measurement and personalisation solutions. For the past few years, we have also been deploying acceleration programs in all business areas (Sales, Supply Chain, Operations, Call Centers, HR and Finance, etc.).

We create value wherever there is data, and work with our clients to improve their processes and create customized business AI applications.

Can you give us concrete examples that show how Artefact designs AI solutions that improve business competitiveness?

Data is the key to understanding customers, developing better products and services, and streamlining internal operations to reduce costs and waste. Artefact supports over 1,000 clients worldwide, including 300 international brands in sectors from consumer goods, retail & e-commerce, and healthcare, to bancassurance, telecoms, industry, energy - and more.

For example, we've been working with the **Orange Telecommunications group** for over six years, and among the many use cases for leveraging the company's automation and AI potential, we deployed a solution with their teams to optimize their technicians' interventions on the fiber network. The solution is based on visual recognition technology that helps operators improve the quality of their installations or repairs. This application, available on a tablet, is currently used by more than 10,000 Orange technicians throughout the country - a resounding success!

ARTEFACT

This case perfectly illustrates Artefact's firm belief that to achieve true data maturity, companies have no choice but to make data accessible to everyone: not only to experts, but also to operational staff in the field. This will lead to new forms of augmented work, where applications and their interfaces put intelligent information in everyone's hands to work more efficiently and with more autonomy.

Artefact also helped the **Carrefour Group** in reducing the carbon impact of its e-commerce branch with a solution that can be implemented by the company and consumers. Carrefour's aim is to become the world leader in food system transformation for all by committing to four major objectives, including achieving carbon neutrality by 2030 for its e-commerce activities. The challenge for Artefact was to enable Carrefour to reliably measure all greenhouse gas emissions from data storage, transport and logistics activities, from first click to final delivery,

Our solution measured greenhouse gas emissions generated by e-commerce orders, then collected activity data to convert it into carbon emissions. All Carrefour business teams helped obtain the data – which is why the operation was a success, as it allowed all stakeholders to become ambassadors for **the group's "carbon neutrality 2030" objective.**

How is Artefact able to always be at the forefront of AI through core research and advanced technology?

At Artefact, we've implemented major projects to ensure that we always leverage **the best of data science and AI technologies** for our clients:

• **The launch of the Artefact Research Center,** which fosters a robust data and AI R&D ecosystem by connecting PhD talent at Artefact with esteemed professors from top universities (Polytechnique, Sorbonne University, and CentraleSupélec, University of Paris-Saclay) and leading enterprises including Orange, Société Générale and Decathlon, with other companies joining us soon. Through the developments and publications of the Artefact Research Center, we aspire to shape a future where AI is not only a powerful tool but is also tailored to the needs of businesses with ethics and responsibility, thereby facilitating its adoption.

• **The creation of the SKAFF technology platform,** an open source developer portal that includes a central software components catalog supporting TechDocs and a scaffolder for automating engineering processes. This platform enhances efficiency by swiftly delivering high-quality outcomes through the consolidation of technical assets, convictions, and tutorials focused on our core technologies.

After a decade of exponential growth, what is Artefact's ambition for the coming years?

First of all, our gratitude goes to our clients for entrusting us, a cornerstone of our success.

I believe that our success also stems from our unique ability to transform data and AI into value for companies. We offer our 1000+ clients **a unique combination of innovation (Art) and data science (Fact).**

By creating multidisciplinary teams and breaking down silos between business and technology departments, we generate real, immediate impact for clients. Artefact has become one of the first and few consolidated pure data & AI players in the market, with the most comprehensive set of data-driven services and AI applications.

We offer data acceleration programs, industry specific AI solutions, and data-driven marketing services. Our engineers build tech agnostic solutions, combining custom code with open source and proprietary software, backed by strong partnerships with leading cloud providers, to create exactly what you need for your data and AI transformation.

Today, Artefact is present in 20 countries across Europe, Asia, the Americas (North & South), the Middle East and Africa, with 23 offices and 1,500 employees. And we have robust plans for geographical expansion as well as an ambitious M&A policy that will continue.

We're also continuously hiring new consulting Partners and Directors, experts in their respective fields, orchestrating collaboration across Artefact's regions. They provide dedicated support and industry-specific services. While strengthening our positions in CPG, Retail, and Luxury, we've also intensified our development in Financial Services, Healthcare & Pharmaceuticals, and Manufacturing, reinforcing human resources.

We're excited about the promising future that AI holds for individuals and organizations. The excellence of AI technology will be realized through the collective capabilities of human talent.

Becoming data-driven (and reaping immediate benefits) has never been easier in the retail sector

Data technologies are now at the head of the aisle in the great supermarket of retail opportunities. Serve yourself: take advantage of these unbeatable offers today!

Jérôme Petit
Managing Partner
Retail & eCommerce
ARTEFACT

Particularly turbulent market conditions in 2022 disrupted the retail sector's pricing and supply chain policies. While price optimization has always been a daily concern for this industry, the 5.2% increase in consumer prices over the whole of 2022 (compared to 1.6% in 2021) and its acceleration in the first quarter of 2023 have intensified the need to react. More frequent price adjustment decisions must be made every day to absorb cost increases or react to competitor repositioning, and be immediately deployed in store networks.

To make matters worse, a shadow still hangs over supply chains, inherited from pandemic events. Many retailers are no longer sure they can offer all products due to lack of stock.

To address both of these challenges, players who are able to rely heavily on their data and process it instantly and at scale through machine learning stand out. Of course, you can't go from data-rich to data-driven overnight, but contrary to common fears, the maturity of the technologies is such that the first building blocks, which can rapidly bear fruit, can be deployed in mere weeks.

Machine learning to optimize prices…

Inflation has focused retailer attention on the daily challenge of price optimization: how to identify products that are most sensitive to price variations (elasticity)? Which products shape the price image of a brand (known value items)? How to react in real time to competitor repositioning? Since promotions represent a growing share of sales (more than ever, consumers are looking for bargains that fit a tight budget), how can they be optimized without impacting profitability? All the answers are in the data! Customer data and proprietary sales history can be combined with external data such as competitor price and promotion records, seasonality, calendar events, and, yes, even the weather (!) to make the right decisions. Retailer pricers and category management teams usually do all this. Unfortunately, they often need to work in Excel but have limited time to devote to it, due to their other responsibilities.

So, what's new? Why, it's the availability today of proven off-the-shelf algorithms that automate the simplest decision-making

processes. It's the seamless and massive availability of third-party data in the market. It's the ease of use of technology stacks that allow millions of transactions to be processed in a few milliseconds. Today, it only takes three months to build a data platform that combines all transaction data, promotions, stock, product hierarchy, store hierarchy, customer data, etc. And a technology partner can manage the infrastructure, resource deployment and network dimensions in the cloud through its managed services. Today, like Monsieur Jourdain, if you know Excel and PowerPoint, you're a data analyst without realizing it: in a matter of days, you can take control of data in BigQuery (Google), Synapse (Microsoft), or Snowflake and build interactive dashboards in Looker, Power BI or Tableau.

…and better manage inventory

In recent years, the health and geopolitical context has also challenged supply chains. Today, the supplier service rate varies from week to week and delivery times can be very uncertain. Distribution channels have also grown highly complex: not only do stores need to be stocked, but home deliveries, click-and-collect, and partnerships

need to be served as well. Once again, data science comes to the rescue of retailers by allowing them to better control inventory management. By leveraging machine learning, retailers can now analyze receipts in real time to immediately detect out-of-stock items, calculate the spread of uncertainties across all links in the chain to better size buffer stocks, or improve stock allocation under an infinite number of constraints (to optimize costs, shorten delivery times, or reduce carbon footprint).

Democratizing data use throughout the company: data is about people

In a business where margins are so tight that operational excellence is a necessity, the notion of a data-driven company is far from new. What's changing today is the ease of access and use of technological platforms.

If technology is no longer a barrier, the challenge is still to make these solutions available to the widest possible audience. To democratize their use, simple solutions need to be deployed on a massive scale, employee training programs need to be multiplied, whether on-demand or more intensive, and events (e.g., hackathons) need to be organized to engage the managers who are driving

the transformation. Data is about people. This is Artefact's slogan, and rightly so.

Directly monetizable data… But perhaps not immediately for everyone

The icing on the cake is that data itself is a goldmine, thanks to retail media and data sharing. As digital signals become more difficult to capture, the billions of transactions and customer interactions that retailers generate have become a critical strategic advantage for them. This data, which provides in-depth understanding of consumer expectations, has great potential for monetization. But it represents a profound, existential transformation of the retailer business model: moving from a self-financed model (with negative working capital) but with very narrow margins, to a model where initial investments are substantial but margins are high. A Copernican revolution, perhaps not the easiest to undertake for all players.

In the great supermarket of retail value creation opportunities, data technologies are now at the top of the shelf, in self-service. Retailers, why wait to share these unbeatable offers with your partners?

CARREFOUR GROUP
How Data & AI can accelerate sustainable business transformation

How Artefact is helping Carrefour achieve carbon neutrality for its e-commerce activity?

Carrefour Group is the leading European retailer and the world's second largest retailer, and is present in more than 30 countries. Carrefour's international profile raises a number of ecological challenges and a desire to offer its customers, regardless of their level of awareness, quality food and services accessible to all.

The Group's aim is to become the world leader in food system transformation for all by committing to four major objectives, including achieving carbon neutrality by 2030 for its e-commerce activities.

To achieve carbon neutrality, three main levers of action have been identified:

Measuring the ecological impact of a delivery in order to manage the strategy;

Reducing the carbon emissions of its logistic and digital infrastructures;

Engaging customers to become participants in ecological transformation.

This aim also provides a triple opportunity for the Group's e-commerce activity: reducing its operating costs, improving its NPS score (customer satisfaction indicator) and anticipating legal changes.

To seize these opportunities and take concrete action on each of these levers, Carrefour must be able to measure all greenhouse gas emissions by drawing on real data that compiles all data storage, transport and logistics activities, from first click to final delivery, whether to the home or by store pickup.

SOLUTION

Defining a reliable, actionable, transparent carbon measure for Carrefour and its customers

The first step for the Artefact and Carrefour teams was to agree on the scope of action for measuring this carbon footprint. They decided to limit themselves to measuring greenhouse gas emissions generated by e-commerce orders in 2021

"*The challenge we gave Artefact was to calculate the CO2 emission of an online order.How much CO2 will a customer produce if their order is delivered or if it's picked up at the store?*"

Bertrand Swiderski
Chief Sustainability Officer
CARREFOUR

The second step was to collect activity data in order to convert it into carbon emissions. As this data wasn't already present and documented in Carrefour's data platform, the business teams (logistics, warehouses, e-commerce) had to be brought together to obtain it. This step proved to be crucial to the operation's success, as it allowed all stakeholders to become ambassadors for the group's "carbon neutrality 2030" objective.

Carrefour's strategy for measuring its carbon footprint was based on a systemic, unifying, long-term, iterative approach. The strategy was successful thanks to the participation of over 30 employees and the involvement of Carrefour customers via their "Engaged Consumers Clubs".

"*Today, we recognize that consumers are becoming experts on these topics. They want to understand how things are done and want to challenge companies. Thanks to them, the project has matured.*"

Léonard Cahon
Consulting Manager - ARTEFACT

Encouraged by these initial results, Carrefour will continue its commitment by publishing the carbon footprint of each of its orders on its e-commerce site in the near future.

"*Soon, customers will clearly see the number of kilograms per CO2 on their orders, thanks to the insights gained from our carbon assessment.*"
Manuel Chatain
E-commerce CSR Manager - CARREFOUR

Opening a wider field of possibilities and ecological alternatives

By analyzing its carbon footprint and implementing this first measure, Carrefour now has a way to pilot its e-commerce carbon emissions reduction strategy. The group can now encourage its customers to review their consumption patterns in order to be more responsible, encourage its service providers to reduce their emissions, and also promote internal awareness by proposing several possible levers of action:
•Act on the choice of delivery slots in order to optimize truck loading, routes and schedules;

•Increase the number of clean vehicles (biogas, electric or hydrogen) by 2030;
•Reduce the amount of packaging used.

And to ensure the sustainability of this measurement and its easy use by the teams, Artefact teams worked on three elements:

•A dashboard to run trajectory simulations by combining forecasted activity data;
•Training to teach how to use and modify the dashboard;
•Detailed documentation to enable employees and clients to understand and reproduce the measurement process from start to finish.

Using data as a key lever to help businesses achieve their environmental objectives

Carrefour's e-commerce, supply chain and logistics platform teams worked together to meet this challenge, supported by the collaboration and expertise of the Artefact, Aktio and Google teams. The project is part of Artefact's "Data for Sustainability" solutions, which aim to create a positive impact on the environment through data by accelerating the ecological transformation of businesses.

"We expected a very clear vision of what each basket order would emit. Our request was complex, but Artefact responded to it with great dynamism and agility."

Bertrand Swiderski
Chief Sustainability Officer - CARREFOUR

"At Artefact, we believe data will play a major role in helping companies achieve their carbon neutrality goals."

Vincent Blaclard
Partner - ARTEFACT

ARTEFACT

Data-driven marketing: the rise of the Customer Data Platform

Florian Thiebaut
Managing Partner - Data-Marketing
ARTEFACT

Everything seems to justify the current explosion of the Customer Data Platform (CDP) market. CDPs' main advantage over older generation Data Management Platforms (DMPs) is that they easily integrate identifiable first-party data (email, phone number) and aren't dependent on using third-party cookies or browsing data to refine customer and prospect knowledge.

CDPs are a true asset in a world that is becoming increasingly cookie- and ad ID-free. At a time when the pandemic is forcing brands to digitise at breakneck speed, and when the transformation of the technical and regulatory environment surrounding advertising trackers is forcing data marketers to revise their approaches, CDPs are here to optimise the customer experience.

A game-changing technical and legal environment

Following Safari's lead in 2016, the world's three main browsers eliminated (or will eliminate) the use of third-party cookies. On the mobile/tablet devices side, Apple's iOS 14 now requires explicit consent for any mobile ID collection.

As for regulation, GDPR laws in Europe have given consumers more control over their personal data, requiring them to give explicit consent for the use of cookies. This regulation represents a major shift in the world of data-driven marketing, as it has reduced the number of cookies placed on European devices by 30%.

This global trend restricting the use of IDs and advertising cookies sharply impacts the targeting capabilities of advertisers, who are often dependent on third party data. The vast majority

of them use or have used retargeting and old generation DMPs that rely heavily on segments fed by third party data.

Along with targeting, measurement must also be transformed. With more stringent consent collection requirements, it's more difficult to collect the consumer IDs needed to track impressions, clicks or views, and reconstruct complete customer journeys.

Four pillars for a sustainable data strategy

To maintain the same performance and differentiate themselves from the competition, advertisers must design a sustainable data strategy and exploit their customer and prospect data to its full potential.

This requires focus on four actions:

The CDP: The first step is to establish a CDP environment based on a suite of tools that is both compliant and sustainable. This will enable data to be collected, stored, processed, visualised and activated, whatever the source. From this foundation, the focus must be on first-party data.

Data governance: Brands need to rethink data governance and processes to enable secure and compliant end-to-end data collection.

Audience segmentation: This data, centralised for a unified view of the consumer, can then be used to create new audience segments and define new metrics for measuring campaign results.

Second-party partnerships: In addition, it's becoming increasingly strategic to form so-called "second party" partnerships with other partner companies to exploit first-party data and create win-win situations.

This data completes a database that is incomplete at certain points in the consumer journey. Examples might be an agreement between an FMCG brand and a retailer, a mobile phone manufacturer with a telco or a hotel chain with an airline.

Three types of data to activate via a suite of tools

First- and second-party data are key to meeting the challenges of the post-cookie world. But what are they and what tools can be used to manage them?

PII or Personally Identifiable Information is essentially CRM (customer relationship management) data. It can precisely identify an individual and is often an email address or a phone number for example. Once anonymised, it can be used via the APIs of media partners (e.g., Google Customer Match, Facebook Custom Audience/conversion API,

Amazon, WeChat, etc.) to build audience segments, perform audience extensions, and reconstruct paths to measure the influence of digital campaigns on offline sales, etc.

Non-PII data can be browsing data that cannot lead directly to the identification of an individual. It can be used to build more granular segments via analytics and audience creation solutions for measuring precision marketing actions without relying on third party data

Data that is purely media-related, such as campaign impressions, video views and click rates, is more voluminous and less granular than the other two types of data. It is more difficult to use but there is a robust market of tools capable of treating it in a secure and compliant manner, such as Google Ads Data Hub, Facebook Advanced Analytics and Amazon Marketing Cloud.

These different data flows are injected into an ecosystem of interconnected tools, which are useful for a range of tasks — from data collection to performance measurement of the actions carried out — and can be activated on all channels, whether media, direct marketing or site personalisation. This entire ecosystem, the result of all the connections built between the different tools already used by the company (also known as "full-stack" solutions), is what is called the CDP.

When it comes to the adoption of this way of working, the numbers don't lie. Fundraising for CDP providers is soaring, the tech giants are all positioned in the sector, and the number of users is exploding.

In fact, according to the Customer Data Institute, the market increased 30% from $1 billion in 2019 to $1.3 billion last year. Estimates see this figure reaching $1.55 billion in 2021 as conditions are even more favourable for the adoption of CDPs.

As the data-driven world continues to evolve at a rapid pace, there seems little doubt in the business value of the CDP. Now is the time for organisations to consider deploying this future-facing technology.

CONFORAMA
AI-enabled personalization boosts Conforama CRM campaign revenues

CHALLENGES

Conforama is the second largest home furnishings retailer in France and is present in seven countries, with 300 stores, including 200 in France. The company sells furniture and decorative items in kit form and posted sales of 1.7 billion euros in 2022.

As a gateway brand, Conforama's goal is to "Make what people want most accessible at the best price." It's an ambition backed by a transformation plan to deliver an omnichannel experience through data and AI. An initial audit and data marketing vision with Artefact identified and prioritized 12 use cases and 25 technical and organizational enablers. The first use case was to integrate a personalized product recommendation into the company's weekly emails.

Several challenges needed to be addressed through this use case:

• How to understand the needs of three million customers and recommend the most relevant products from a catalog with 42,000 references?

• How to propose only products currently in stock, on promotion, and not already suggested to customers?

• How to easily operate and maintain the technical solution?

Conforama

> "*Time savings, yes, but above all a business benefit for our CRM teams. Because thanks to this personalization, customers click more and therefore buy more. We've gained 15% of the click rate following the personalization of these emails, which represents several million in incremental sales.*"

Mélodie Charles, *Marketing Director*
CONFORAMA

SOLUTION

Saving consumers time, improving business productivity

By using machine learning algorithms to analyze user data, such as preferences, purchase history and online behavior, artificial intelligence-based product recommendation suggests products relevant to consumers in a personalized way. This allows companies to better understand their customers' needs and recommend products that match their interests, resulting in increased sales and customer retention.

One of the main benefits of this solution is that it saves customers time. Rather than scrolling through countless product pages to find what they're looking for, customers can quickly access a selection of recommended products that specifically meet their needs. AI-based product recommendation can enhance the online shopping experience and encourage customers to return for more purchases. A strategic advantage, given that 72% of consumers only interact with marketing messages that are personalized and tailored to their interests.

In addition, AI-based product recommendation can boost business productivity: machine learning algorithms can analyze large amounts of data in real time, allowing companies to continuously monitor customer trends and buying behavior. This can help organizations better understand customer desires and quickly adapt their product offerings accordingly. It can also enable companies to optimize their inventory by offering products that are more likely to sell, which can lower costs and maximize profits.

Lastly, AI-based product recommendation can offer significant business benefits. By suggesting relevant and personalized products to customers, companies can improve their conversion rate, increase sales and strengthen their brand image. From a market perspective, AI-based product recommendation has been shown to deliver +2.5% incremental growth.

A first use case focused on email campaign personalization

Conforama

Prior to this project, all Conforama customers received emails featuring the same eight products selected each week by the marketing teams. This was a labor-intensive task, as it required identifying the eight products most likely to interest three million customers, each of whom had unique interests. All this time spent analyzing data could have been spent on more strategic activities, such as creating editorial content for those emails.

Today, an email is sent to every Conforama customer each Tuesday containing eight product recommendations. But these recommendations are personalized according to purchase history, and filtered exclusively for products that are on sale, are available in stores, and that haven't been featured in previous activations.

The implemented AI solution includes 4 main data processing steps:

• Collection of transaction histories, customer and product references, then data preparation;

• Building the "Collaborative Filtering" model to calculate customer appetite for the product catalog;

• Product filtering based on available inventory, commercial news (sales, promotions, etc.), past activations and purchases;

• Product data enrichment (photos, prices, descriptions, etc.) for activation.

This solution is based on 16 data tables, 25 transformation and modeling steps, and 40 automated quality tests. Dozens of iterations of the model made it possible to choose the most efficient approach based on transaction history. Thanks to this solution, Conforama now generates several million recommendations each week in 45 minutes at a cost of 50 euros per week.

In other words, if you count development and operation costs, as well as incremental sales, the project break-even point is reached in one week, with an automated and reliable solution.

Conforama

A smooth transition to AI: lessons from Conforama's success story

For many players, there are three challenges linked to their level of maturity:

LEVEL 1

Personalizing a currently rule-based touchpoint using an AI algorithmic approach;

LEVEL 2

Extending AI-based personalized recommendation across the entire customer journey (similar products / complementary products / suggestion based on purchase history);

LEVEL 3

Optimizing the orchestration of recommendations across channels to ensure an omnichannel experience.

Level 1 is often the most difficult, as it requires laying the foundations for four separate dimensions: target vision, user experience and priorities; data sources; technological tools; project team and work method.

The Conforama example offers valuable lessons about these four dimensions:

• Select a first use case and functionalities that can be quickly implemented and measured to put the organization on the road to success. For example, this initial victory means Conforama can now plan the deployment of product recommendations in stores or the improvement of their algorithm thanks to browsing data.

• Ensure the data is reliable. Good data modeling relies first and foremost on good quality data. For Conforama, exploratory analyses were performed on more than 50 tables to select data sources in areas such as customer knowledge, product repositories and transactions.

• Use technologies that allow teams to deploy a technical solution quickly and collaboratively. Conforama selected the most appropriate tools for this type of workflow: DBT, BigQuery ML and Vertex AI for their performance, modularity and portability.

• Build a dedicated team capable of dealing with all potential problems, and adopt a test and learn approach. To do this, a multidisciplinary IT / Conforama business team was formed, and a 2-week sprint approach was adopted.

Retail Media:
An indispensable asset for brands

While Retail Media represented only 9% of digital media investments for brands in 2019, it will soar to 43% of these investments in 2023 and is expected to double in value by 2024 to reach €100bn. Vincent Cailliot, Director of Data Consulting and Sidney Zeder, Senior Consulting Manager – Data Marketing, both of Artefact, explore the opportunities of retail media for Consumer Packaged Goods (CPG) brands.

Vincent Cailliot
Director Data Consulting
ARTEFACT

Sidney Zeder
Senior Consulting Manager
ARTEFACT

Retail Media investment driven by regulatory and tech developments

The increased importance of Retail Media in the digital strategies of brands can be explained by the evolution of the availability of consumer data, which is at the heart of any media personalization strategy.

In the past, third-party cookies were mainly used to drive this strategy. Regulatory developments to better protect consumer data, including the GDPR in Europe, have led to new technological constraints, making cookies obsolete: This process began as early as 2016, when Safari removed cookies from its platform. It continues today with Google Chrome, the most popular browser, announcing the removal of third-party cookies from its platform next year.

As a result, brands need to find new data sources to build their digital activation strategy. One solution is for brands to better leverage their first-party (1P) data assets, i.e., their proprietary data, by collecting more and better data from their customers. But to create digital activation strategies with long-term value, proprietary data is no longer sufficient: it needs to be enriched with data – especially transactional data – from retail partners.

Facilitated data sharing between brands and retailers allows brands to deploy ever more relevant digital marketing strategies with a high level of autonomy. This increased collaboration and data sharing between retailers and brands is possible thanks to "data clean rooms" such as Amazon Marketing Cloud, LiveRamp or Decentriq, which allow

the sharing of personally identifiable consumer data at the individual level in an anonymous way.

A rapidly evolving ecosystem

The ecosystem of technology partners around Retail Media is highly fragmented and constantly evolving, with partners that are more or less specialized depending on major Retail Media activities: first-party, second-party or third-party cookie data collection tools, data processing and audience creation, activation or analysis, etc. The challenge for brands will be to identify which combination of technology partners will best meet their needs, depending on their current technical ecosystem and their own business challenges.

Retailers are also an essential part of this ecosystem, providing access to transactional data to build their Retail Media strategy. While the majority of retailers in the US have launched Retail Media offerings, most retailers in Europe are still in the experimentation and use case-testing phases; few have yet industrialized use cases with brands.

Valuable use cases beyond Retail Media

LeRetail Media allows brands to address marketing use cases from consumer insight generation to digital campaign activation and marketing performance measurement. The availability of transactional data (previously unavailable to CPG B2B2C brands) at the "individual" level enables the construction of insights and activation plans that are all the more impactful. The same data can be used to measure their effect on sales and ROI, enabling effective optimization of activation plans.

Retail Media is just the next step towards more collaboration between retailers and brands. In a long-term partnership perspective, collaboration and data sharing can enable the implementation of more advanced category management and supply chain use cases, such as the analysis of the long-term value of existing promotions or the prediction of in-store product demand and thusoptimize supply

chain operations.

Which Go-To-Market strategy to launch?

For retailers, it's important to define a new offer to monetize their data. This can range from monetizing their owned media inventory (website), to sharing data "as a service" in a clean room, to offering services (campaign management or reporting as proposed by Amazon for example). These new offers can be marketed internally or via partners. The internal or external development strategy will determine the associated costs, in terms of salaried resources (commercial and technical profiles to be recruited) and technical resources (clean room tools, technical infrastructure to be set up).

For brands, the approach will be more traditional, from defining their business needs by identifying and prioritizing use cases, to setting up the partnership with their key business partners. Likewise, the implementation of pilots with a retailer to test the value of use cases can be carried out by a third-party partner.

UNILEVER

How does Artefact support Unilever on Retail Media use cases to increase its sales?

CHALLENGES

Thanks to Retail Media, Unilever identifies new growth opportunities and increases sales of its priority product categories.

The global Unilever group has a portfolio of 400 brands that are anchored in the daily lives of its 5 billion consumers. Present in France for 125 years, Unilever is the leader in several market categories including ice cream, condiments and deodorants.

The consumer packaged goods (CPG) sector has the particularity of being intermediated. Since the distribution of their products is carried out by different retailers, there is little direct relationship between CPG brands and their clients. Yet, consumer knowledge is key to optimize media and promotional strategies, product assortment in the territory or to identify new growth levers.

It is from this challenge that the so-called Retail Media

offers were born, i.e. 2nd party data partnerships. A retailer's data is made available to a brand to enrich its own data assets in a win/win approach. This movement was initiated in 2012/14 by pure players such as Amazon and Alibaba, and traditional players such as Casino, Walmart and Carrefour have gradually followed.

Accompanied by Artefact since 2019 on various data issues, the Unilever Group seized this opportunity to identify new growth levers, develop a common consumer and product approach with a leading retailer and increase sales of certain priority product categories. This is the case, in particular, for the Magnum brand's mini ice cream bars range.

"Retail Media is a win/win strategy for brands and retailers. Retailers' data allows us to enrich the shopper's knowledge and accurately measure our activities on all channels, throughout the transformation tunnel. For their part, retailers find a new source of revenue and differentiation from their competitors. In addition, it is a way to better satisfy their clients with more personalized offers and a better anticipation of stock shortages."

Sarah Baqa
Head of Performance Marketing - Unilever

Retail Media in a win/win partnership with Trade Marketing

Advertising expenditure (media and traditional marketing) in the consumer goods sector amounts to nearly 680 billion dollars worldwide and 16 billion in France. The investments allocated to Trade Marketing are just as massive.

While the advertising levers have been highly digitized over the last twenty years, Trade Marketing is still very underdeveloped. However, driven by e-commerce, it represents half of the budgets of CPG advertisers (600 billion dollars worldwide, 16 billion in France). Trade marketing is strategic for these brands because it consists in carrying out actions in partnership with mass retailers to better meet consumers' expectations: prospectus, merchandising, point of sale animations but also sales forecast and supply chain optimization. Thus, the digitalization of trade marketing represents a real growth opportunity for advertisers and brands.

Retail Media is based on this concept of partnership in the promotion of brands' products. The sharing of 2nd party data favors client knowledge and allows a better segmentation and therefore a better targeting of advertisements or promotions. This targeting can be activated in both proprietary brand environments and in external audience extension environments.

Retail Media infuses Unilever's media, promotional and supply chain strategy

Unilever, together with Artefact, has identified 6 strategic axes to exploit the full potential of Retail Media:

- **Media activation**
- **Measuring Customer Lifetime Value**
- **Coupon customization**
- **Optimization of store assortments**
- **Short-term sales forecast**
- **Supply of shelves**

Since the CPG group initiated these retail partnerships, each of these axes has been worked on through concrete use cases. For this, Unilever has benefited from Artefact's end-to-end support. This concerns data strategy, the launch of pilots, the construction of new data products, but also the provision of human resources (data analysts, data scientists and data engineers), as well as the training of Unilever employees.

SOLUTION

An iterative approach

The Retail Media project was envisaged through an iterative approach in 4 steps:

- **Identify and select the right distribution partners** based on their pre-existing relationship with Unilever and their technology infrastructure.

- **Laying the foundations of the partnership**, i.e. the use cases that can be implemented as well as the necessary legal and technical requirements. The modus operandi is also built to facilitate collaboration between all stakeholders (agencies, internal and external teams of Unilever and the brand).

- **Implement pilots for each use case.** They allow for testing and optimization of devices on key Unilever brands (Magnum and Sun) before scaling up.

- **Deploy devices on a greater number of brands**, use cases, markets and partner brands.

The vision of this data partnership is therefore based on a virtuous circle: the measurement of actions carried out makes it possible to collect insights which in turn feed the next activations.

"Our data strategy is really ambitious, so we wanted to be accompanied by a team of experts. We chose Artefact, which already has experience with similar products and works in a very advanced way on data and Retail Media solutions. In addition, they have the technical capabilities and resources to take use cases to scale. This is the key to this long-term partnership."

Sarah Baqa

"Retail Media allows us to activate audiences and precisely measure the link between digital actions (advertising, promotions) and sales. The data from the retailer's loyalty cards also has the advantage of tracking consumers over time (Customer Lifetime Value) to see if they have increased their purchase frequency, if they have increased the value of their basket or if they have switched brands, etc."

Florian Thiebaut,
Managing Partner – Data Marketing - Artefact

RESULTS

A use case conducted on Magnum's mini bars segment

For example, retailers' data was activated in the media to meet a challenge for the Magnum brand: to confirm its leadership during the summer period, which is key for this product line, and to acquire new clients in the Confectionery segment (mini bars).

To achieve this, the first step was to use shopper data to identify the different consumer segments. We showed them the same banners during a defined period. This allowed us to measure the increase in on and offline sales, but also to track the buyers recruited to retain them post-campaign. Through this first wave, Unilever was able to identify the best performing and most profitable audiences. The ultimate goal is to use these results at scale to target these audiences with more personalized ad creative and promotional offers.

This pilot campaign significantly increased the brand's revenue.

Unilever's data and Retail Media strategy will continue in 2022 alongside Artefact. In addition to the two «Precision Media» levers, the exploitation of the 4 other identified work axes is also planned.

Data monetization opportunities for retailers: Retail Media within the CPG/Retailer data ecosystem

Retail media has been around for quite a while, but thanks to the evolution of new uses of consumer data, its potential is gaining attention. Sidney Zeder, Senior Manager and Gaétan Bélan, Senior Data Consultant & Product Owner, both of Artefact, explore the opportunities of data monetization for retailers.

Gaétan Bélan
Senior Data Consultant
& Product Owner
ARTEFACT

Sidney Zeder
Senior Consulting Manager
ARTEFACT

Retail media has been on the rise on digital platforms for the last six years, most notably on Amazon. The Covid-19 crisis accelerated this trend for traditional retailers. Retail media, in simple terms, is the means for retailers to sell media inventory on their e-commerce platforms. Because the Covid-19 pandemic fueled the shift to digital ways of buying, such as e-commerce or click-and-collect, even for grocery shopping, retailers had no choice but to go with the flow.

In fact, between 2019 and 2020, CPG e-commerce penetration increased by five points, from 10 to 15 percent. For retailers, the downside is that margins are lower in e-commerce than in brick-and-mortar. The upside is that by selling online, they collect a lot of consumer data that can be monetized or used to create new services. In a Goldman Sachs study, 82% of CPG companies surveyed said they were already investing in at least one retail media platform. This represents approximately 17% of digital budgets already allocated to retail media.

Media investment has effectively shifted down the marketing funnel. Although many brands' search investments still flow into the "Google family," we're seeing brands diversify their digital spend into e-commerce platforms to capitalize on the "search destination" status they hold. When you're on Amazon as a consumer, you're very close to the "moment of truth": you're in a purchasing mindset. Therefore, when you're on Amazon as a brand or product, the closer you can get to that funnel, the better. Goldman Sachs expects this trend to translate into a 6-8 percent increase in total CPG e-commerce sales through retail media over the next four years.

Retailer data monetization opportunities with CPG brands

This close-to-the-funnel media investment trend has created opportunities for retailers around three types of data monetization with CPG brands:

1. Inventory monetization: traditional retail media consisting of selling media inventory on proprietary assets. This can be offline inventory – retailers have long monetized their customers to offer coupons or specific promotions to brands in their stores – but also their online inventory on their own platforms, such as their e-commerce website, where brands can display banner ads, emails or even shopping mobile applications to deliver personalized promotions to their customers.

2. Data monetization: retailers are monetizing existing consumer data across CPG brands to support their customer centricity. 1P data shared by retailers originates from their loyalty program. The cardholder data they share can be socio-demographic (e.g. the age of their consumers), transactional (e.g. what did they buy),

behavioral (e.g. what did they look at), loyalty data (e.g. did they buy again), etc. This data is shared "as-a-service" in a data clean room where brands can access the retailer's data in a secure environment to carry out specific use cases defined by the two partners.

Carrefour, for example, has created a consumer intelligence service called Carrefour Links, based on the LiveRamp clean room, where partners can access their cardholder data. This is a self-service platform that allows users to perform basic activities such as reconciling retailer and brand databases on individual customers to build a more complete view of the consumer and thus improve their experience. It also provides analytics and measurement capabilities that Carrefour can bill to its partners.

Access to this data can unlock three types of use cases for brands:

•Marketing: the data shared by retailers allows brands to gain insights about their consumers, activate them with media, or measure marketing performance through transactional data. For example, an ice cream brand partnered with a

retailer to build advanced audiences for a digital marketing campaign. Using the retailer's transactional data, the brand was able to build and activate two audiences: the brand's current ice cream buyers and ice cream buyers from competitor brands. As a result, the brand was able to increase the uplift of its campaigns by targeting the two relevant audiences with adapted messaging.

•Trade: the data shared by retailers allows brands to perform revenue growth management use cases by better optimizing promotions or assortment… It also unlocks store optimization use cases through enhanced in-store experience or sales force optimization. For example, one brand worked with a retailer to analyze the short- and long-term impact of promotions on incremental margin. This enabled them to identify certain types of promotions that were margin-destroying for both the brand and the retailer, as opposed to those that generated a positive long-term business impact.

•Operations: the data shared by retailers allows brands to optimize

their supply chain through demand forecasting and demand management use cases. It can also fuel sustainability as well as production and innovation use cases.

Brands usually start implementing marketing use cases to deliver short-term value with a simple set-up, while long-term partnerships can then address very valuable use cases for Trade and operations that benefit both partners. Across all three categories, data monetization unlocks better measurement capabilities: optimized media performances, customer lifetime value calculation, dynamic budget allocation and global ROI optimization.

3. Service monetization: along with inventory and data monetization, retailers can unlock additional revenue by providing different levels of services to brands.

In the most developed additional service offering, retailers can propose managed services to brands, based on shared SLA and KPIs. In a long-term partnership approach, retailers optimize their revenue potential through best-in-class services for brands. Amazon offers its key clients advanced services such as category management studies and dashboards or MMM (Media Mix Modeling) studies to help them improve their strategy in relation to this retailer.Des cas d'usages à valeur, au delà du retail media

Data monetization: maturity in the Retail Media market

Artefact has benchmarked over 20 retailers in the US Retail Media market and analyzed their maturity with respect to Retail Media based on their value proposition and the comprehensiveness of their capabilities. We found that most retailers in the US have launched a Retail Media offering but are still in a nascent stage, mainly offering

inventory to brands, while more mature retailers are focusing on data monetization or even service monetization for the best-in-class retailers.

Data monetization offerings require setting up technical capabilities, such as a technical infrastructure to collect, store and process the first-party data to be shared, as well as a clean room to share data with brands, or even partnerships with DSPs to enable brands to directly activate audiences created in clean rooms. But the business opportunity is worth it: selling data as-a-service delivers margins often in excess of 80%, compared to inventory monetization where margins typically only reach 40%, as inventory assets are limited and therefore less scalable.
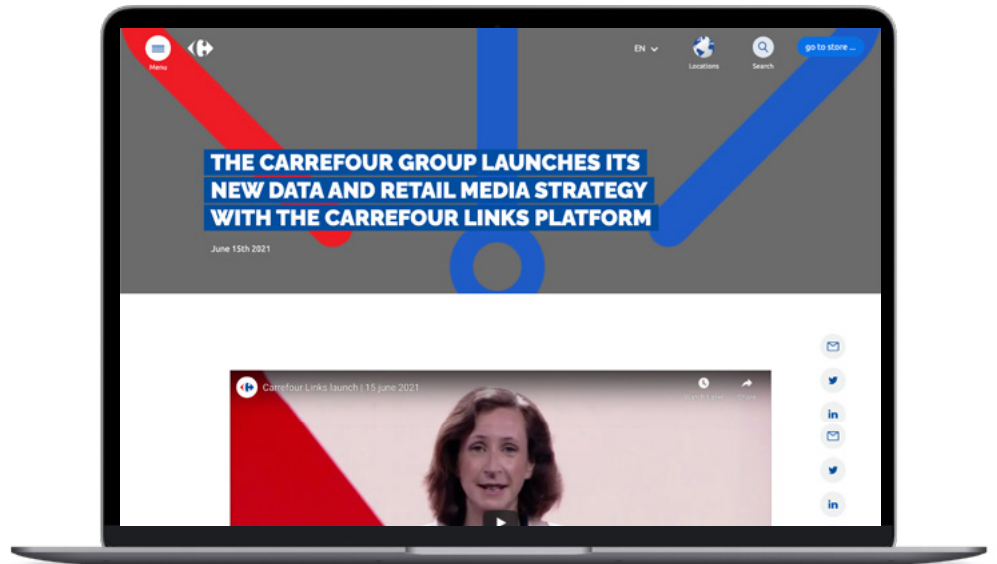
CARREFOUR LINKS

With Artefact are helping brands to improve media targeting thanks to Carrefour's data

CHALLENGES

Retail ecosystem is currently going through a revolution

The old rules of retail have been deeply impacted by two new factors: digital assets (digital media) and shopper data, that helps leverage knowledge of customer behavior and granular activation of digital assets.

Retail media mixes digital assets and shopper data. For many years, retail media has been helping brands and manufacturers become more efficient in four different ways:

- **customer insights:** a deeper understanding of your customers thanks to data from identified members of customer loyalty programs

- **precise targeting:** moving from mass marketing towards 1:1 marketing

- **activation:** omnichannel activations that can be synced through several touchpoints enhanced by shopper data

- **measuring:** measure of the return on ad spend and impact on sales

SOLUTION

Carrefour's vision on their Retail Media system

Carrefour provides manufacturers with cutting-edge solutions to conduct retail media campaigns and to develop their sales and display presence both in store and on their ecommerce website.

According to Nicolas Trolé, Carrefour Media has the unique opportunity to be able to leverage assets from the whole entire division, such as 5000 physical stores, a digital platform that reaches 10 millions customers every month, and the key element in their omnichannel offering, their unique data.

In 2021, Carrefour took the decision to develop a new full stack platform on carrefour.fr, with Criteo as partner, to offer advertisers new activation solutions and better measuring and targeting.

"When we talk about data at Carrefour, we are talking about data from our customer loyalty program, which represents 14 million cardholders, 7 million customers registered online and offline, and 5 million digital-only customers"

Nicolas Trolé,
Chief Revenue Officer - Carrefour Links

*"Retail media is an ecosystem that needs a certain number of mandatories on the retailer's and manufacturer's sides"*

Thomas Faure,
*Senior Consulting Manager
E-Retail Lead - Artefact*

RESULTS

Artefact as a trusted partner to maximise the value of Carrefour's data

Thomas Faure reviews the four pillars of the partnership between Carrefour Links and Artefact:

• **Business consultancy:** Manufacturers want to understand where growth comes from. "Carrefour's data helps brands identify business opportunities for them to accelerate"

• **Profiling and customer analysis:** "Who is my customer?" One of the main challenges for manufacturers is to define who their customer is. Carrefour's data helps brands represent the personas of their brand buyers

• **Strategy, operation and activation:** helping brands create media activation strategies (budget, KPIs)

• **Advanced data use cases:** projects that take into account the whole retail value proposition
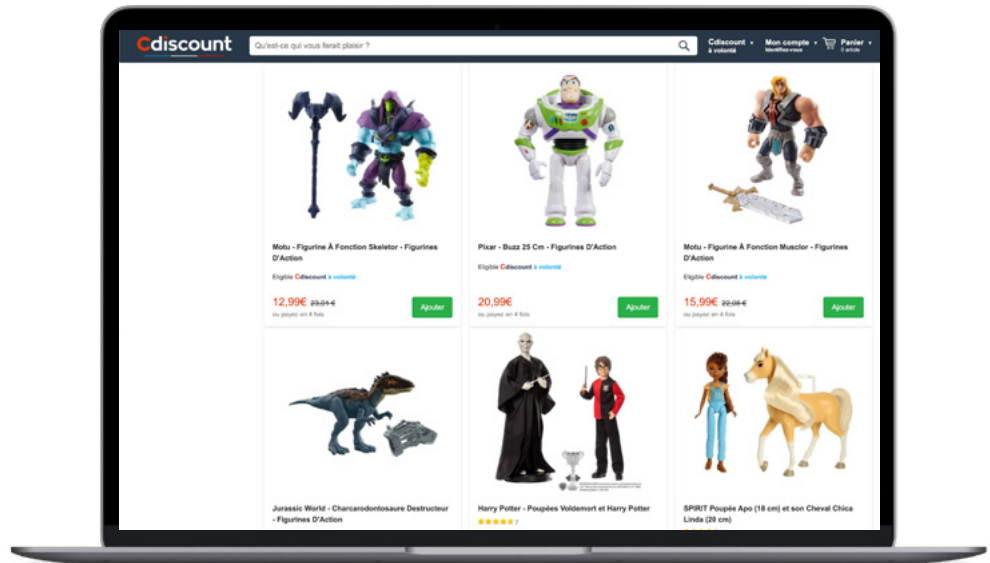
ARTEFACT

How Artefact helped boost Mattel's online sales on Cdiscount (RelevanC Advertising) retail media platform?

CHALLENGES

Increase presence and visibility for the brands of the Mattel Group on e-retailers media ecosystems by leveraging retail media.

E-retail is a strategic distribution channel for the Mattel Group as **32% of transactions** in the gaming category were conducted online during the 1st semester of 2020.

Retail media is used in an objective of performance enabled by leveraging shopper data. Retail media is especially helpful for the conversion part of the sales funnel.

Retail media lives in different types of environments such as:

- **in-platform**, products that maximise the presence of products directly on the eCommerce website such as sponsored brands, sponsored products, merchandising banners

- **off-platform**, solutions that generate additional traffic redirected to the product pages on the eCommerce website such as search engine ads, social ads, display & video ads

High touch handling of Mattel's retail media investments in partnership with relevanC Advertising.

In 2020, Cdiscount cemented its position as the number one French e-retailer, generating 25 million unique visitors and 10 million customers (adding a million new buyers during the year).

Cdiscount's vision of its retail media ecosystem covers the entire sales funnel and encompasses different steps (brand awareness, consideration, traffic, acquisition, and insights)

In February 2020, relevanC introduced a self-serve platform, relevanC Advertising Platform, built entirely in-house, that offers retail media solutions through search and display on Cdiscount.

"relevanC Advertising is a best-in-class tool to operate and manage retail media campaigns,"

says Maïana Darmendrail – *Digital Manager & E-Retail Manager, Mattel France*

"CDiscount possesses very mature retail media solutions in search, display and video,"

states Thomas Faure – *E-Retail Lead, Artefact*

Artefact's made the decision to perform an all-year round online campaign aimed towards buyer intent. Artefact defined custom audiences out of precise shopper data items (purchase history, buyer intent, search history, browsing history, income level ..).

The key to increase performance of media investments was to conduct continuous optimizations on both retail and media KPIs.

"We have the intimate conviction that continuous optimization brings performance improvements,"

explains Cédric Chamoux – *Directeur Retail Media, relevanC Advertising*

- Optimizations were made on **several factors: channels** (onsite, offsite, search, display), **segments** (audience, keywords), **creatives, formats**.

- Optimizations were based on **several metrics**: **media metrics**, **business metrics**, **retail metrics** (stock level, promotions, organic positioning)

- Optimizations happened on different solutions: **search advertising**, daily improvements on **keyword selection**, **display advertising**, bi-weekly improvements on **impact measuremen**t

In addition to leveraging the e-retailer ecosystem, Artefact kept looking for ways to innovate and test new solutions to build more expertise and scale projects.

In that way, Artefact conducted off-platform campaigns through Shopping for Partners solutions (*Google for Retail*) that were activated to support milestones, such as product launches.

RESULTS

Ad campaigns six times more effective!

Mattel achieved extremely positive results through retail media activations in 2020:

- **600% increase** on return on ad spend for media investments using retail media data (compared to traditional campaigns)

- **from 4 up to 10 euros** earned when spending one euro on display and search ad campaigns

"relevanC Advertising has the most qualitative data that you can find on the market. The daily optimization by Artefact really moved the needle,"

mentions Maiana Darmendrail – *Digital Manager & E-Retail Manager, Mattel France*

Brands are able to find success when mixing business metrics with media expertise, i.e. sync retail signals (stock level, promotion level) with media KPIs to better optimize advertising campaigns.

CARREFOUR GROUP

How to reduce food waste in the bakery-pastry department?

CHALLENGES

How to reduce food waste in the Bakery-Pastry department?

Predicting demand as accurately as possible is one of the foundations of the retail business. However, this challenge is becoming more complex as consumer habits evolve. From now on, it is necessary to take into account the combined use of different sales channels (from e-commerce, to local stores, to supermarkets) and the increasing demand for social and environmental responsibility. At the top of the list of CSR issues linked to mass retailing: food waste.

In France, bread is the third most wasted product after fruits and vegetables. Indeed, bread, pastries and cakes are fresh products that have a very limited shelf life. But, they are also items that can generate a lot of frustration among customers when they are out of stock at the end of the day. This is why Carrefour Group, together with Artefact, sought to use machine learning and data science to optimize the prediction of sales in the bakery-pastry department of its supermarkets.

This project meets a double objective: produce enough to meet the demand, while reducing "scrappage", i.e. the volume of unsold products.

This use case's success relies, in particular, on the teams involving both the "jobs" concerned (department managers) and the various technical profiles of Artefact and Carrefour Group.

SOLUTION

A new algorithm to prevent "scrappage" in the bakery-pastry department

In the retail world, every day is different. Sales are highly dependent on the context: holidays, weather, current promotions, merchandising highlights, etc. To take into account all the variables that have implications on demand, we need to be able to analyze the petabytes of data generated by the billions of transactions carried out every year at Carrefour, to which we need to add external data that influences consumption. These calculations are only possible with Artificial Intelligence.

The Carrefour Group and Artefact teams thus started with data from the sales receipts generated by more than 200 supermarkets in France. Every day, this data is collected, cleaned and enriched with external sources – such as calendar data, for example – to build a sales history over several years. This represents thousands of configurations for a single day, depending on the assortment, product prices, promotions, etc.

This data is used to train supervised machine learning models, built on the basis of decision trees, which determine the relationships between the target variable (future sales) for each product and the explanatory variables (promotion, cannibalization, etc.).

Close collaboration with the "field" teams

From an organizational point of view, the project was led by multidisciplinary teams. Two teams on the Artefact and Carrefour sides combined technical and business profiles. The operational skills of Carrefour's retail professionals played a crucial role. They were able to explain their business, their needs, and bring their vision, in order to guarantee the success and adoption of the solution "in real life".

For example, the algorithm's recommendations were first tested in pilot stores, only on pastries, to get feedback from the field teams. Their comments were used to improve the models, before the solution was deployed across all bakery and pastry products in the supermarkets.

Easy-to-access and interoperable tools

All of Carrefour's data-driven use cases are fed by a centralized platform in the cloud, which makes the data accessible, formatted and documented. The results of the processing are then fed back into Carrefour's information systems, shared by a wide range of employees. Several teams, made up of data scientists, data engineers, but also data translators (pivotal profiles, acting as a link between business and data), are likely to consult, transform and process them for specific uses.

RESULTS

Concrete results on the number of discarded products...and on the brand image

A few months after the implementation of this new prediction model, the results are very positive.

In fact, over the last five months of 2021, approximately 100 tons of pastries were saved. At the same time, sales have increased due to fewer shortages at the end of the day.

Finally, the Net Promoter Score, a performance indicator closely monitored by Carrefour, has evolved very positively.

Carrefour Group is multiplying AI use cases to improve customer experience

For the group, future experiments follow the same model: responding to business needs, working jointly with operational teams, to feed the customer experience. This acceleration of Carrefour's digital transformation was made possible by the creation of complete and expert data teams within the company, and the deployment of data platforms in all countries where the group operates.

The volume and wealth of data collected by Carrefour provides a unique opportunity to explore the major challenges facing the retail sector: omnichannel, e-commerce, anticipation of consumer habits, etc. Carrefour recently unveiled other examples of how data can be used to improve the customer experience: five-minute shopping on Carrefour.fr, the implementation of personalized assortments for local stores, and the personalization of promotions.

In fact, over the last five months of 2021, approximately 100 tons of pastries were saved. At the same time, sales have increased due to fewer shortages at the end of the day.

Demand forecasting: Using Machine Learning to predict sales in retail.

Jérôme Petit
Managing Partner
Retail & eCommerce
ARTEFACT

All industries aim to manufacture just the right amount of products at the right time. But, for retailers, this issue is even more important as they also need to manage their stocks efficiently. Too many items in stock is bad. Too few items in stock is also bad. And to predict sales as close to possible, retailers used to only rely on the previous years past sales. This method is useful only to a certain point and suffers many biases. Thankfully, Machine Learning has now evolved to be able to provide very accurate predictive models using different signals based on how they influence purchases.

Retailers main challenge is to go beyond past sales to predict future sales accurately

Managing orders and inventory is the one of the strongest competitive advantages that can help retailers achieve success. And it is a real challenge to master as it involves processing a huge number of SKUs - some of them that are even perishable - ordered daily. We estimate that bad inventory management, whether it's out of order items or excess stock, cost US retailers close to two billion dollars per year. For decades, retailers have been relying on the analysis of their past sales for their Enterprise Resource Planning (ERP) that helps them reduce their investment and exploitation costs. However, these methods are heavily biased and are not that useful when trying to predict accurate sales.

Dozens of signals to take into account when assessing stock levels

The reason why predicting sales appears to be so complex and difficult is because, in a given period, many factors can affect purchase: weather, purchasing trends, regulation, product launches, global pandemic, buying behaviors ... And the main issue with these types of predictions based on past recordings is that they don't factor individual incidents, and they make monthly sales appear as if they were perfectly distributed when they were probably not. In fact, an out of order item might have caused a slowing of sales of a particular product or a particular category, but it won't show in the monthly reports. Even worse, bad numbers are viewed as a mark of buyers disinterest, when they could be the opposite as consumers over-purchase an item and cause it to sell out. It is also important to note that a missing product in store doesn't

necessarily mean that the product is out of stock. Big box retailers struggle to restock their shelves in real time so a product that becomes instantly popular might disappear from the shelves and thus perform worse than expected, when in fact, it is available in inventory. Retailers are in need of technology that can help them step into a new paradigm that could seamlessly align offer and demand.

Using Machine Learning to help employees in stores

Machine Learning is the solution to these challenges. Predictive models are now able to forecast sales up to four weeks by using a number of signals that can affect sales, such as the season, the current trends or the price levels. The models are based on three indicators, the Day, the Product, and the Store. It is important that they don't rely on a single indicator to be the most accurate possible. For the sake of the argument, let's imagine that you analyse the season signal. Your data is going to be biased because predictions based on dates are not 100% helpful because a certain date can fall on a weekday in a given year and fall on a weekend in the following year, and sales vary enormously between the weekday and the weekend. Also, this date could end up on a holiday, or a calendar event (such as Christmas, Thanksgiving, Easter) or a sports event. All factors that might increase the consumption of some products. As another example, let's have a look at the price levels signal. Promotions at store level can severely affect the sales of a product from a given category or even raise the attractiveness of a store as a whole. Retailers need to be able to predict the impact of their promotion strategy. That's why it is important to take into account many different signals and indicators to accurately forecast sales through Machine Learning and advanced Artificial Intelligence models.

Concrete use cases of Machine Learning on inventory management

The technology is there. Now for retailers to use it effectively and make accurate predictions, they need to be able to collect and analyse a huge amount of data. The problem lies in the fact that they have different sources of data and it can become complex to try to process multiple Excel and PDF files that contain previous reports and media plans. Retailers need to adopt big data tools that can process this information into a clean and lisible digest, that is necessary to be able to create predictive models that can prevent inventory issues.

The vast majority of retailers are sitting on years of old sales data. However, this data can turn out to be inaccurate because of the effect of past promotions or events (heat wave, sports event, major local event). To get rid of this bias, predictive models are able to combine past sales numbers with those of similar stores. Artefact has enjoyed success with this approach and conducted a successful experimentation in China on the 020 platform that raised by 20% the level of accuracy in sales predictions.

The other big challenge is to be able to prevent items from being unavailable in shelves when they are available in inventory. It is almost impossible for store managers to ask their employees to monitor shelves in real time and restock them immediately. Technology solutions using surveillance cameras and weight sensors are a huge investment. However, available information such as real-time sales, the characteristics of a SKU and the organization of a particular shelf can be useful when predicting out of order. Models are able to analyze the usual flow of sales of an item, meaning the time between two sales of the same product in a given store. In case of a statistical anomaly, a human is notified for an intervention in the store to analyze the situation and solve it.

There are many ways in which Machine Learning can help traditional retailers. Predictive analysis is only one of them. Retailers can gain a lot from relying on advanced technology to increase their store revenue by better managing their inventory. But processing vast amounts of data can also help them on the supply chain front, or on the merchandising side. Well-devised tools can become a real asset to managers by taking care of complex and time-consuming tasks and provide them with accurate reports in a short amount of time and letting them shift their priorities to improve customer experience.

Sales forecasting in retail:
what we learned from the M5 competition.

In this article, Data Scientist Maxime Lutel sums up his learnings from the M5 sales forecasting competition, which consisted in predicting future sales in several Walmart stores. He will walk you through our solution and discuss what machine learning model worked the best for this task.

Maxime Lutel
Data Scientist
ARTEFACT

Using machine learning to solve retailers' business challenges

Accurate sales forecasting is critical for retail companies to produce the required quantity at the right time. But even if avoiding waste and shortage is one of their main concerns, retailers still have a lot of room for improvement. At least, that's what people working at Walmart think, as they launched an open data science challenge in March 2020— the M5 competition — to see how they could enhance their forecasting models.

The competition aimed at predicting future sales at the product level, based on historical data. More than 5000 teams of data lovers and forecasting experts have discussed for months about the methods, features and models that would work best to address this well-known machine learning problem. These debates highlighted some recurring issues encountered in almost all forecasting projects. And even more importantly, they brought out a wide variety of approaches to tackle them.

This article aspires to summarize some key insights that emerged from the challenge. At Artefact, we believe

in learning by doing, so we decided to take a shot and code our own solution to illustrate it. Now let's go through the whole forecasting pipeline and stop along the way to understand what worked and what failed.

Problem statement: Hierarchical times series forecasting

The dataset contains 5-year historical sales, from 2011 to 2016, for various products and stores. Some additional information is provided, such as sell prices and calendar events. Data is hierarchically organized: stores are divided into 3 states, and products are grouped by categories and sub-

categories. *(cf Image 1)*

Our task is to predict sales for all products in each store, on the days right after the available dataset. It means that 30 490 forecasts have to be made for each day in the prediction horizon

This hierarchy will guide our modeling choices, because interactions within product categories or stores contain very useful information for prediction purposes. Indeed, items in the same stores and categories might have similar sales evolution, or on the contrary they could cannibalize each other. Therefore, we are going to describe each sample by features that capture these interactions,

Image 1
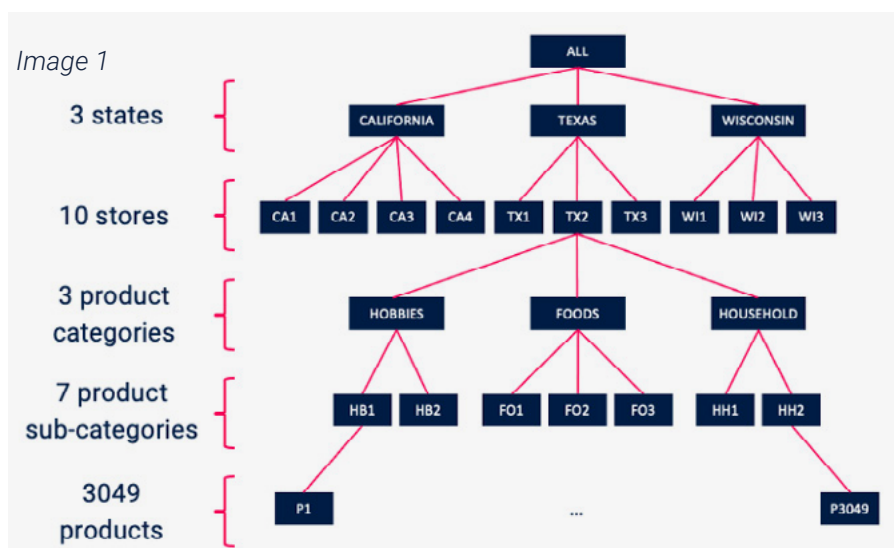
and prioritize machine learning based approaches over traditional forecasting ones, to consider this information when training the model.

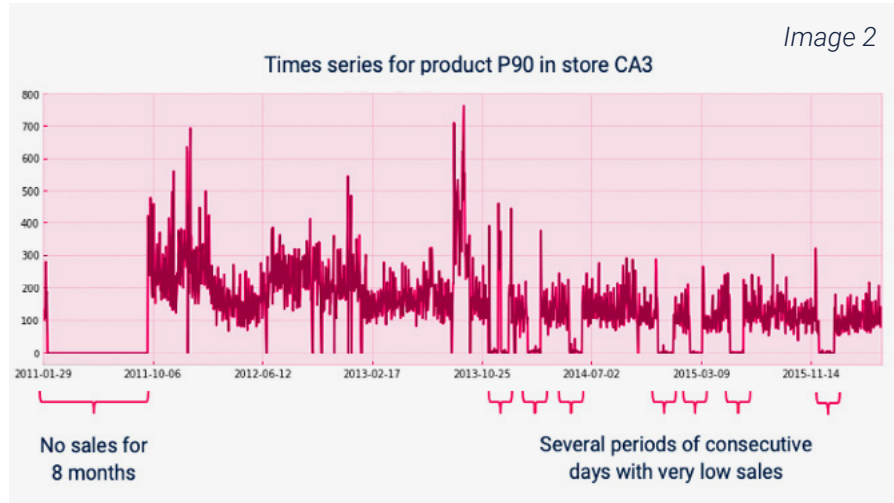## Two main challenges: intermittent values and an extended prediction horizon

At this stage, you might think that it is a really common forecasting problem. You're right and that's why it is interesting: it can relate to a wide range of other projects, even if each industry has its own characteristics. However, this challenge has 2 important specificities that will make the task more difficult than expected.

The first one is that the time series we are working with have a lot of intermittent values, i.e. long periods of consecutive days with no sales, as illustrated on the plot below. This could be due to stock-outs or limited shelves' area in stores. In any case, this complicates the task, since the error will skyrocket if sales are predicted at a regular level while the product is out of shelves. *(cf image 2)*

The second one comes from the task itself, and more precisely from the size of the prediction horizon. Competitors are required to generate forecasts not only for the next week, but for a 4-week period. Would you rather rely on the weather forecast for the next day or for 1 month from now? The same goes for sales forecasting: an extended prediction horizon makes the problem more complex as uncertainty increases with time.

Feature engineering — Modeling sales' driving factors

Now that we have understood the task at hand, we can start to compute features modeling all phenomenons that might affect sales evolution. The objective here is to describe each triplet Day x Product x Store by a set of indicators that capture the effects of factors such as seasonality, trends or pricing.

Image 2

Times series for product P90 in store CA3

No sales for 8 months

Several periods of consecutive days with very low sales

SEASONALITY

Rather than using the sales date directly as a predictor, it is usually more relevant to decompose it into several features to characterize seasonality: year, month, week number, day of the week… The latter is particularly insightful because the problem has a strong weekly periodicity: sales volumes are bigger on the weekends, when people spend more time in supermarkets.

Calendar events such as holidays or NBA finals also have a strong seasonal impact. One feature has been created for each event, with the following values:

• Negative values for the 15 days before the event (-15 to -1)
• 0 on the D-day
• Positive values for the 15 days following the event (1 to 15)
• No value on periods more than 15 days away from the event
The idea is to model the seasonal impact not only on the D-day, but also before and after. For example, a product that will be offered a lot as a Christmas present will experience a sales peak on the days before and a drop right after.

TRENDS

Recent trends also provide useful information on future sales and are modeled thanks to lag features. A lag is the value of the target variable shifted by a certain period. For any specific item in a given store, the 1-week lag value would be the sales made one week ago for this particular item and store. Different shift values can be considered, and the average of several lags is computed as well, to get more robust predictors. Lags can also be calculated on aggregated sales to capture more global trends, for example at the store level or at the product category level.

PRICING

A product's price can change from one store to another, and even from one week to another within the same store. These variations strongly influence sales and should therefore be described by some features. Rather than absolute prices, relative price differences between relevant products are more likely to explain sales evolutions. That's why the following predictors have been computed:

• Relative difference between the current price of an item and its historical average price, to highlight promotional offers' impact.

• Price relative difference with the same item sold in other stores, to understand whether or not the store has an attractive price.

• Price relative difference with ►►

other items sold in the same store and same product category, to capture some cannibalization effects.

Categorical variables encoding

Categorical variables such as the state, the store, the product name or its category also hold a significant predictive power. This information has to be encoded into features to help the model leveraging the dataset hierarchy. One-hot encoding is not an option here because some of these categorical variables have a very high cardinality (3049 distinct products). Instead, we have used an ordered target encoding, which means that each observation is encoded by the average sales of past observations having the same categorical value. The dataset is ordered by time for this task to avoid data leakage.

All categorical variables and some of their combinations have been encoded with this method. This results in very informative features, the best one being the encoding of product and store combination. If you wish to experiment other encoders, you can find a wide range of methods here.

Tweedie loss to handle intermittent values

Different possible strategies can be used to deal with the intermittent values issue. Some participants decided to create 2 separate models: one to predict whether or not the product will be available on a specific day, and a second one to forecast
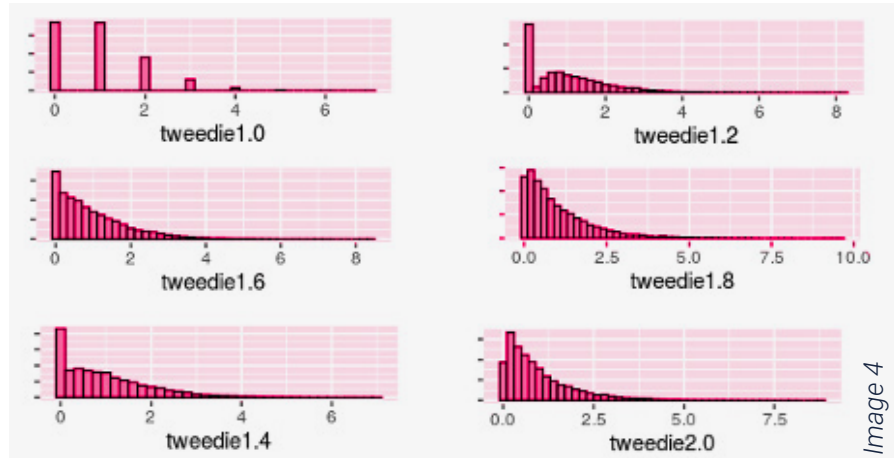
Image 4

sales. Like many others, we have chosen another option, which is to rely on an objective function adapted to the problem: the tweedie loss.

Without going into the mathematical details, let's try to understand why this loss function is appropriate for our problem, by comparing sales distribution in the training data and the tweedie distribution *(cf image 3)*.

They look quite similar and both have values concentrated around 0. Setting the tweedie loss as an objective function will basically force the model to maximize the likelihood of that distribution and thus predict the right amount of 0s. Besides, this loss function comes with a parameter — whose values are ranging from 1 to 2 — that can be tuned to fit the distribution of the problem at hand *(cf image 4)*

Based on our dataset distribution, we can expect the optimal value to be between 1 and 1.5, but to be more precise we will tune that parameter later with cross-validation. This

objective function is also available for other gradient boosting models such as XGBoost or CatBoost, so it's definitely worth trying if you're dealing with intermittent values.

How to forecast 28 days in advance?: Making the most out of lag features

As explained above, lag features are sales shifted by a given period of time. Thus, their values depend on where you stand in the forecasting horizon. The sales made on a particular day D can be considered as a 1-day lag if you're predicting one day ahead, or as a 28-day lag if you're predicting 28 days ahead. The following diagram illustrates this point *(cf image 5)*

This concept is important to understand what features will be available at prediction time. Here, we are on day D and we would like to forecast sales for the next 28 days. If we want to use the same model — and thus the same features — to make predictions for the whole forecasting horizon, we can only use lags that are available to predict all days between D+1 and D+28. This means that if we use the 1-day lag feature to train the model, that variable will also have to be filled for predictions at D+2, D+3, … and D+28, whereas it refers to dates in the future.

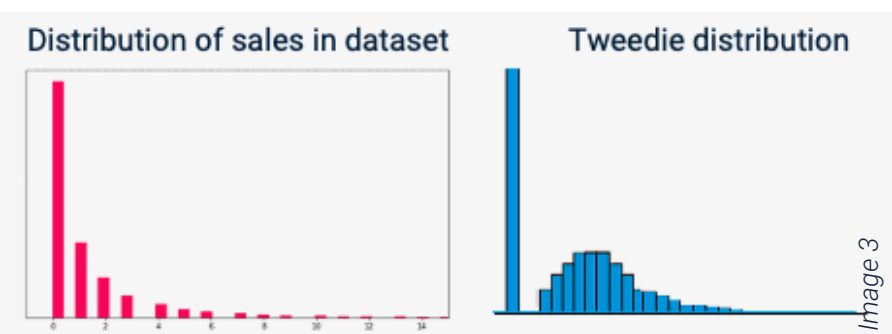Still, lags are probably the features with the biggest predictive power, so
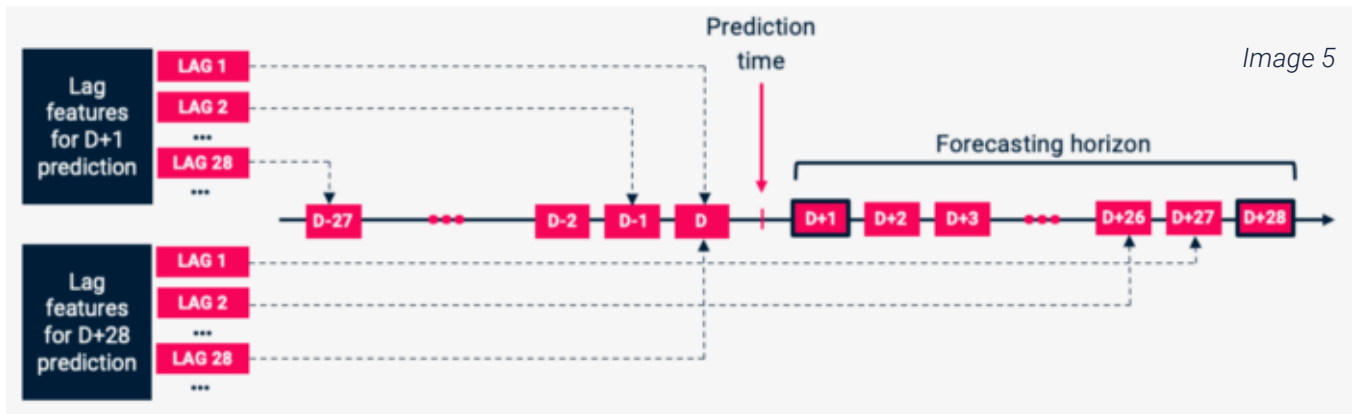

Image 3

Image 5

it's important to find a way to make the most out of this information. We have considered 3 options to get around this problem, let's see how they performed.

OPTION 1:
ONE MODEL FOR ALL WEEKS

The first option is the most obvious one. It consists in using the same model to make predictions for all weeks in the forecasting horizon. As we just explained, it comes with a huge constraint: only features available for predicting at D+28 can be used. Therefore, we have to get rid of all the information given by the 27 most recent lags. It is a shame as the most recent lags are also the most informative ones, so we have considered another option.

OPTION 2:
WEEKLY MODELS

This alternative consists in training a different LightGBM model for each week. On the diagram above, every model is learning from the most recent possible lags with respect to the constraint imposed by its prediction horizon. Following the same logic as the previous option, it means that each model can leverage all lags except those that are newer than the farthest day to predict.

This method allows us to better capitalize on lag information for the first 3 weeks and thus improved our
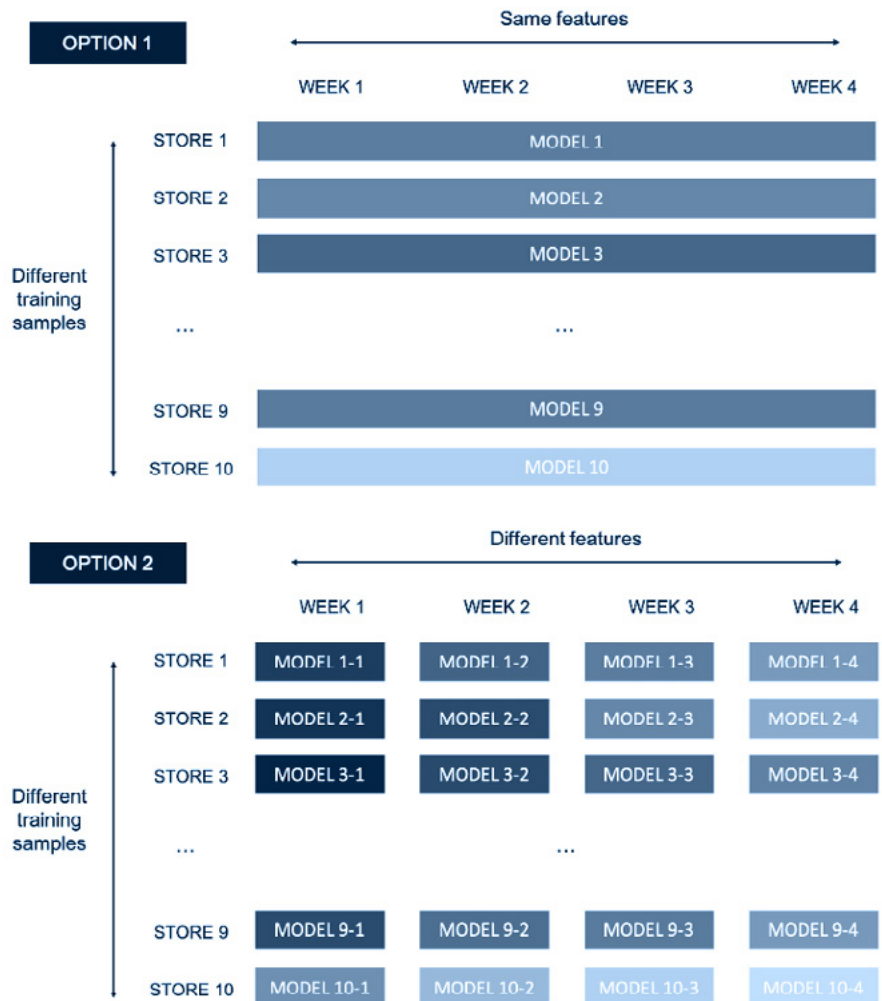
MORE PRECISELY:

- Model 1 makes forecasts for days 1–7, relying on all lags except the 6 most recent ones.

- Model 2 makes forecasts for days 8–14, relying on all lags except the 13 most recent ones.

- Model 3 makes forecasts for days 15–21, relying on all lags except the 20 most recent ones.

- Model 4 makes forecasts for days 22–28, relying on all lags except the 27 most recent ones just like in option 1.

►►►

solution's forecast accuracy. It was worth it because it was a Kaggle competition, but for an industrialized project, questions of complexity, maintenance and interpretability should also come into consideration.

Indeed, this option could be computationally expensive and if we are aiming at a rollout on a whole country scale, it would require maintaining hundreds of models in live. In that case, it would be necessary to evaluate if the performance increment is large enough to justify this more complex implementation.

OPTION 3:
RECURSIVE MODELING

The last option also uses weekly models, but this time in a recursive way. Recursive modeling means that predictions generated for a given week will be used as lag features for the following weeks. This happens sequentially: we first make forecasts for the first week by using all lags except the 6 most recent ones. Then, we predict week 2 by using our previous predictions as 1-week lags, instead of excluding more lags like in option 2. By repeating the same process, we always get recent lags available, even for weeks 3 and 4, which allows us to leverage this information to train the models.

This method is worth testing, but keep in mind that it is quite unstable as errors spread from week to week. If the first week model makes important errors, these errors will be taken as the truth by the next model, which will then inevitably be poorly performing, and so on. That's why we decided to stick with option 2, that seems to be more reliable.

Ensuring model robustness with an appropriate cross-validation: Why cross-validation is critical for time series

In any machine learning project, adopting an appropriate cross-validation strategy is critical to
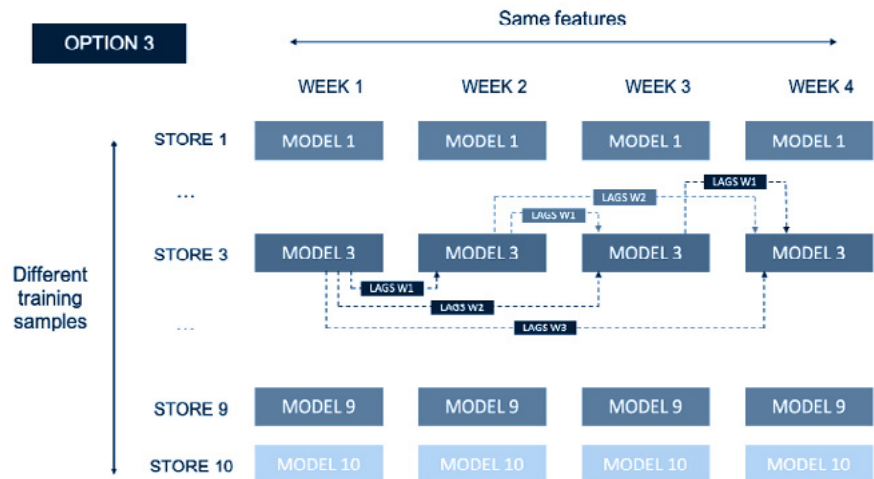
simulate correctly out-of-sample accuracy, select hyper-parameters thoroughly and avoid over-fitting. When it comes to forecasting, this has to be done carefully because there is a temporal dependency between observations that must be preserved. In other words, we want to prevent the model from looking into the future when we train it.

The validation period during which the model is tested also has a greater importance when dealing with time series. Model performance and the optimal set of hyper-parameters can vary a lot depending on the period over which the model is trained and tested. Therefore, our objective is to find which parameters are the most likely to maximize performance not over a random period, but over the period that we want to forecast, i.e. the next 4 weeks.

Adapting the validation process to the problem at hand

To achieve that goal, we have selected 5 validation sets that were relevant to the prediction period. The diagram below shows how they are distributed over time. For each cross-validation fold, the model is trained with various combinations of parameters on the

training set and evaluated on the validation set using the root mean squared error. *(cf Image 9)*

Folds 1, 2 and 3 aim at identifying parameters that would have maximized performance over recent periods, basically over the last 3 months. The problem is that these 3 months might have different specificities than the upcoming period that we are willing to forecast. For example, let's imagine that stores launched a huge promotional season over the last few months, and that it just stopped today.

These promotions would probably impact the model's behavior, but it would be risky to rely only on these recent periods to tune it because this is not representative of what is going to happen next.

To mitigate this risk, we have also included folds 4 and 5, which correspond to the forecasting period respectively shifted by 1 and 2 years. These periods are likely to be similar because the problem has a strong yearly seasonality, which is often true in retail. In case we had a different periodicity, we could choose any cross-validation strategy that has more business sense. In the end, we have selected the hyper-parameters'

Image 9

STEPS	INCREMENTAL FORECAST ACCURACY
Feature Engineering (seasonality, trends, pricing)	+ 3 pts
One model per store	+ 0.3 pts
Tweedie Loss	+ 0.5 pts
Weekly modeling	+ 0.7 pts
Cross-validated hyper-parameters optimization	+ 1 pts

combination with the lowest error over the 5 folds to train the final model.

Results

The different techniques mentioned above allowed us to reach a 0.59 weighted RMSSE — the metric used on Kaggle — which is equivalent to a weighted forecast accuracy of 82.8%. The chart below sums up the incremental performance generated by each step.

These figures are indicative: the incremental accuracy also depends on the order in which each step is implemented.

Key takeaways

We have learned a lot from this challenge thanks to participants' shared insights and we hope it gave you food for thoughts as well. Here are our key takeaways:

- Work on a small but representative subset of data to iterate quickly.
- Be super careful about data leakage in the feature engineering process: make sure that all the features you compute will be available at prediction time.
- Select a model architecture that allows you to leverage lags as much as possible, but also keep in mind complexity considerations if you're willing to go to production.
- Set-up a cross-validation strategy adapted to your business problem to evaluate correctly your experiments' performance.

ARTEFACT

AI IS ABOUT PEOPLE