ARTEFACT

DATA & AI
TRANSFORMATION
FOR BUSINESS

ARTEFACT

AI IS ABOUT PEOPLE

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

New York
Los Angeles
Mexico
Colombia
Brazil

UK
The Netherlands
Germany
France
Switzerland
Spain
Morocco
Senegal
Lebanon
Saudi Arabia
Dubai
India
South Africa

Chengdu
South Korea
Shanghai
Malaysia
Singapore

21
COUNTRIES

1500
EMPLOYEES

+1000
CLIENTS

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

DATA READINESS | AI ACCELERATION | DATA & DIGITAL MARKETING | TECHNOLOGIES

Data & AI transformation for Business

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.

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

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 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.

Data Readiness

Entreprise Governance in the data age

Vincent Luciani
Co-founder and CEO
ARTEFACT

First, an anecdote about the pitfalls of not using data.

In 1984, a young Michael Jordan – today considered to be the best basketball player of all time – had just won the American college basketball championship in North Carolina. A shooting guard, he was one of the top three picks of the NBA draft, which selects the best American college players. But at the time, shooting guards were considered smaller and less impressive than pivots. Because of this reasoning, and regardless of the data that indicated his extraordinary potential, Adidas twice refused to sponsor Jordan.

The result? Nike – Adidas' biggest competitor – approached Jordan to design his own line of shoes. To date, 100 million pairs of Air Jordans have been sold. The huge commercial success of the Jordan brand still generates a turnover of 3 billion dollars today thanks to the man who was six times voted best player in the NBA. That's a lesson every business should take to heart: always make the most of your data.

How data enables better decision making

Data helps you understand the past: data analysis offers a clearer way of seeing the root causes of problems in a multifactorial world.

Data helps you predict the future: data can forecast consumer demand. A good example is the way we helped Carrefour Group predict sales in their bakery and pastry department. The goal was to sell more by avoiding stockouts, and to sell better by reducing food waste. The idea was to provide the managers of the fresh produce department, who make bread and pastries on site every day, with an accurate estimate of volumes. We used a technique called machine learning, based on learning from historical data. It worked very well: we improved turnover by a few points by avoiding stockouts, while dividing waste by three.

Data helps you optimise what already exists: UPS software gives each of its drivers specific routes to follow, but they aren't always the shortest routes: they don't factor in the distance to be covered, but rather the fewest left turns to be made on each route. By analysing the data, they realised that 60% of all accidents were caused by taking left turns, and only 3% by taking right turns (and requiring more waiting time).

Analysis, prediction and optimisation: with these, data becomes a 'production factor' rather than an 'innovation factor'. All these initiatives are linked to significant ROI, whose stakes can be in hundreds of points of turnover or incremental margins.

Data enables better decisions that improve business performance, but how can the company's decision-making body and governance be adapted to take better advantage of the data age?

Data must be treated as a strategic asset

Companies need to resolve their 'data debt' – a technology debt accumulated around the lifecycle of data. We have a multiplicity of very complex IT systems, which have been superimposed over time, with data sources that are very often poorly documented, sometimes inconsistent, difficult to access, and which don't comply with the rules in force (internal or external such as GDPR). This 'debt' wastes a lot of time in mobilising reliable information for analysis.

A good illustration of this is the case of one of Artefact's major pharmaceutical clients. They were unable, until recently, to obtain their turnover by product figures due to an inability to cross-reference their production plant product repository with their transactional repository, which contains financial data organised by sales outlet: an irrational situation where the company's own data was unavailable for it to use.

The longer data debt is allowed to accumulate, the more costly it becomes

to rectify. Treating data as a strategic asset means agreeing to invest in a program to improve data quality, documentation and accessibility, and to do so in a sustainable manner as sources multiply.

A data driven company must become a talent development factory

Talent has become the decisive factor in the digital age. Access to technology is a commodity so universally accessible that the emergence of no-code, for example, and cloud computing, are increasingly associated with turnkey services such as database storage and operation or automatic algorithm building.

The recruitment war is very serious – among the seven million available job offers in the US posted on LinkedIn (70% of the total) for example, two out of seven are for data-related positions. The pace of technological change is so rapid that it's impossible to establish a competency framework at any given time.

This acceleration is being driven by GAFA-backed big budget cloud technology frameworks, research labs and free algorithms from a global network of 100k researchers, plus the almost immediate adoption and widespread use of open source in the start-up world.

In an environment where technology is changing so rapidly, companies must be able to build their own talent. For large companies in particular, extensive training/retraining programmes will be needed:

• The World Economic Forum (WEF) predicted that by 2025, 85 million jobs will disappear and 97 million new ones involving data will be created.

• AT&T now invests around $250 million annually in T University, which enables existing employees to develop in-demand expertise in areas such as data science and cybersecurity.

When considering the talents that will be needed tomorrow, it's tempting to focus only on technology and assume the next generation will be exclusively composed of engineers and data scientists. Clearly, there will be a need for them and there will even be a shortage of them for the next few years, but this is only part of the story. In a world where data and algorithms can automate manual, repetitive and time-consuming tasks, and where technology is ever more accessible, there's plenty of room for other types of talents: problem-solving, creative, interpersonal, etc.

In a modern company, the decision-making process must be decentralised

In technology, there is a major progressive movement towards decentralisation, which began with technological breakthroughs (e.g., crypto currency or the metaverse which are decentralised systems), but also IT systems (cloud, where we share our machines, or distributed computing architectures such as Hadoop, a world-renowned framework for distributing calculations on different servers).

Decentralisation is also valid in governance. Why? Because centralisation is impossible: there's simply too much data, with poorly controlled sources, which can easily be poorly interpreted without contextual knowledge. Some benefits of decentralisation in governance include:

1. Rapid decision-making and less time spent going back and forth

2. Letting the 'one who knows' make the best decision

3. Empowering decision-makers with a mandate – with limits and a control loop, of course.

This implies a deep organisational change oriented around knowledge: the organisation as an autonomous whole, constituted of cognisant communities organised around knowledge. At Artefact, for example, we made sure that certain entities (chapters, tribes, guilds) could decide on their own very critical things, like salaries, bonuses, prices and offers and even staffing! We have created a fully decentralised governance.

all share them too. They are invaluable tools which break any strategy down into measurable objectives, then into two or three sub-objectives shared by all employees. Andy Grove, who taught them to John Doerr; he in turn wrote the book, 'Measure What Matters', about the process. OKRs allow employees to be valued for their accomplishments, not merely their backgrounds, degrees, or titles.

Conclusion: A new perception of data and data governance

Data changes the way companies are governed, and the role of managers – and directors in particular – from one of making the best decisions in the company's interest to one of creating a system so that everyone contributes to making the best decisions in the company.

As a Board member, you have an important role to play in addressing issues that have a tangible impact on the results of the company, but it will require paradigm shifts. And it is precisely because data shakes up governance that it must be appropriated by governance itself.

As your data governance policy matures, you will need to ensure that your priorities encompass responsible and ethical data processing and energy sobriety initiatives within its framework. It's crucial to operate in ways that reduce the economic, social and environmental footprint of digital technology.

In times of doubt or radical changes, the human tendency is to shut down and turn inward. People build self-defence systems (confirmation bias) to avoid being shaken in their convictions, by creating their own systems of truth. Properly used, data is the truth that should allow us to make better, more reliable and independent decisions for the benefit of customers, users, employees and the company as a whole.

Leading is measuring

Some measurements are unnecessary in decision-making and can easily be dispensed with in favour of other, more proactive management tools. Here's an example to demonstrate this concept:

A Profit & Loss statement is nothing more than the company's report card. It's useless in decision-making and doesn't serve the investor relationship at all. (We almost never look at turnover and EBITDA which is like looking at past performance, not future performance.)

Instead, for Artefact, we spent time constructing a very advanced data-driven finance department. We built a financial data warehouse connected in real time to a data platform, with our key Salesforce tools for pipeline estimation, a staffing tool, our HR database, and our ERP for customer contracts. Out of this, we've created a real-time file containing seven key indicators summarising the operational health of the company looking forward and shared with investors – a tool far more meaningful and helpful for better decision making.

For example, we can respond to and engage with our shareholders on issues such as what our expected demand will be in six months, or what the composition of our workforce is supposed to be (critical in a context of rapid growth and challenging recruitment).

KPIs are broken down into Objectives and Key Results (OKRs) shared among the top partners of the company, and we

ORANGE FRANCE
AI solution of visual recognition at the service of Orange France technical intervention quality

CHALLENGES

AI-based application lets fiber installers easily verify intervention compliance

Orange, a leading telecommunications company in France and Europe, undertook an Artificial Intelligence-based transformation in 2020. This new vision for data and AI aims to help the thousands of technicians, engineers and other professions employed by the group with tasks that are too complex for humans to manage alone. At the same time, an overhaul of data product governing regulations was also launched.

To implement their new strategy, Orange was accompanied by data service consulting firm Artefact. Together, the two businesses industrialized numerous use cases to support the group's Business Units, facilitate technical decision-making and transform the business by realizing the potential of data and AI.

Among these jointly constructed use cases is an AI solution designed to assist Orange technicians in connecting clients to the fiber network. This solution, integrated into the technician's application, verifies that none of their interventions on network equipment generate "defects" or "non-conformities", which are often the cause of the growing degradation of the fiber network in France.

In 2022, we increased the number of connections by 23% compared to 2021. This sustained growth has led to an increase in reported malfunctions. This is why ARCEP made it mandatory, in summer 2021, to take pictures before and after each intervention.These images have a triple objective: to monitor, to intervene as quickly as possible in the event of a problem and to penalize the operators found at fault.

"For Orange, this regulation requires the analysis of 20,000 photos daily. A task impossible to perform quickly and faultlessly without AI assistance."

Médéric Chomel, *VP Data, AI & Automation*, ORANGE FRANCE

An AI solution based on image recognition

Asking technicians to verify numerous control points at the job site, or having enough human resources dedicated to analyzing the 20,000 pictures generated every day is not a feasible solution. This would be too time-consuming, too costly, and would not be error-proof. In addition, sampling is not an option, as each and every intervention must be verified.

To review these tens of thousands of photos daily (10,000 interventions x 2 – 10 photos), Orange and Artefact developed an algorithmic model using image recognition (computer vision). Technicians, via their mobile app, send their photos to an artificial intelligence engine which checks in near real-time whether their work is in conformity. If the technician disagrees with the machine's recommendations, they are free to ignore them. The AI is perfectly integrated into the technician's workflow.
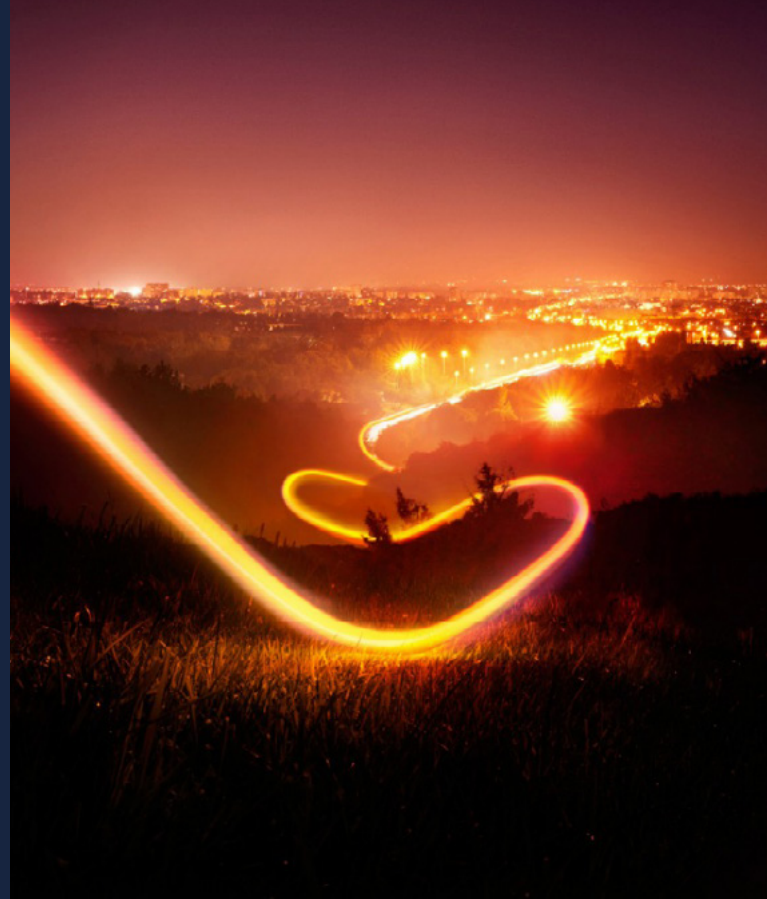
Success factors: quality data, multidisciplinary teams, transfer learning model

The project was led by a multidisciplinary team with a mix of profiles from both Orange and Artefact. A feature team was created, composed of the Product Owner, data scientists, engineers, users, and experts from other professions to work on delivering the solution.

Time was the first issue faced by the team in charge of the project. Orange only had nine months to deploy the first version of their solution. This is why they decided to base part of the project on "transfer learning", a method of using pre-existing models either already in use within the company, or available as open source. Artefact teams then reworked these models via retraining, labeling and pre-processing, to shorten delivery time, and also developed several others from scratch. The team then looked at the response times of different computer vision solutions. Some solutions on the market processed images in seven or even eight minutes, while the target time was three seconds. The application should be launched when the technician is about to leave the site. It is impractical to ask the person to wait 10 minutes to check the conformity of

his installation. To reduce latency as much as possible, calculations are parallelized. Thus, several models are executed at the same time to obtain results in quasi real time.

The third challenge, but certainly not the least, was analysis precision. In order to ensure maximum accuracy, the algorithm had to be supplied with a huge quantity of compliant and non-compliant photos. The labeling of these images was carried out by a partner company called Isahit, which was able to process 80,000 photos in three months of development while respecting data confidentiality.

"This project is part of Orange's long-term AI transformation strategy. We have packaged the code so that it can be reused in future use cases where image recognition is needed. This AI product has already been reused to support fiber technicians in another field of operations."

Vincent Luciani, *co-founder and CEO* of the ARTEFACT group.

Change management: encouraging end-user adoption of the application

To understand their way of working, technicians have been part of the project from the very beginning . This allowed the development team to identify several points, one of which is crucial: the application should not be perceived as a means of controlling the work of technicians, but as a tool to facilitate their daily work.

So, to ensure that end-users are comfortable with the application and that it was ethically designed, the team worked on two aspects.

First, technicians must be able to maintain control over the machine and go against its recommendations. This is why the explainability of the results returned by the model was a core value. If the model finds one or more non-conformities, the AI must specify which area or areas are affected.

Then, once a first version of the application was ready, the team had it tested by 50 volunteer technicians. This allowed the team to collect relevant feedback so they could improve the models. As an example, the conditions in which the photos are taken can lead to confusion between orange-colored cables (from Orange) and red-colored cables (competitors). The recurrence of this error led the feature team to improve the algorithm's acceptability. The model's performance was reduced in order to avoid contradicting what the human sees.

For Vincent Luciani, co-founder and CEO of the Artefact group,

"All of our AI projects are designed to respect the seven fundamental principles for ethical AI use established by a group of European Commission experts. The first of these values is human control. We have placed technicians at the heart of the project to ensure that this new solution makes their daily lives easier and doesn't hinder their autonomy. This has also been crucial for its adoption by all Orange installers."

A success which is part of a global transformation strategy through AI

In just nine months, this new application was designed, tested, corrected and industrialized on a large scale. It was a real technical and human feat, as the tool has now been adopted by the 10,000 technicians deployed every day in France.

This application is just one of the 150 use cases developed by Orange over the last two years as part of its transformation through AI. Since then, Orange – together with Artefact – has put 15 new models into production to support other functions, such as sales or customer service.

Médéric Chomel
VP Data & AI Automatisation
ORANGE FRANCE

"Regulations on artificial intelligence are in the process of being developed. Our transformation, using data and AI, is intended to be respectful of privacy, to benefit humans and their environment, and to be unbiased. This is why this strategy anticipates future regulatory changes as far as possible. We must remember that this type of project is not just technical; humans are the greatest factor in their success."

ARTEFACT

A Practical Approach to Business Impact from Data & AI

Ten Proven Tactics from the Battlefield

Oussama Ahmad
*Data Consulting Partner,
Global Travel & Tourism Lead*
ARTEFACT

Karim Haye
*Data Consulting
Senior Manager*
ARTEFACT

1 - Building data solutions should be driven by the business, for the business

In today's rapidly-evolving business landscape, data intelligence has become an essential tool for companies looking to remain competitive. Organizations that fail to adopt data-driven models risk falling behind their rivals by missing out on valuable insights and opportunities for expansion, optimization and innovation. In short, leveraging data intelligence for business is no longer a luxury, but a necessity for sustainability and evolution, and business leaders should be the ones spearheading the identification, prioritization, and development of data & AI solutions. Contrary to the common belief that business stakeholders are just "consumers" of data solutions, we believe that they should lead the entire process , supported by data and technology experts.

2 - Identifying the "right" data solutions requires in-depth analysis of the business value chain and business processes

A thorough analysis of the business value chain and key business processes is best carried out by business stakeholders themselves. This analysis identifies areas where data solutions can drive significant business impact in the form of revenue growth, cost optimization, customer experience enhancement or operational excellence. During this process, it's essential to identify business opportunities that align with the company's overall business strategy. For example, analysis of the supply chain and its key ratios can help identify potential gaps and inefficiencies that can benefit from data analytics and intelligence.

3 - Prioritizing a few data solutions will ultimately have the most business impact

The goal shouldn't be to impress with a long list of data solutions, but rather to identify the most critical business areas that can benefit from data-driven insights. By avoiding the temptation to pursue too many data solutions, organizations can stay focused and increase their chances of building successful data solutions. It's also important to identify the value-added capabilities of data solutions beyond simple reporting. While reporting is valuable in providing a summary of business performance, it only provides a retrospective view of data, leaving little room for analysis and decision-making. To fully leverage the power of data, organizations must identify data solutions that provide diagnostic analytics that automatically identify the root causes of performance and predictive analytics that anticipate future trends.

4 - Assessing feasibility of data solutions requires a full understanding of data sources and technologies

Before embarking on the development of a data solution, it is vital to conduct a detailed feasibility study that examines the availability and quality of the required data sources, as well as the cost of the technologies and expertise required to collect and process these data sources. This includes examining the hardware and software requirements, as well as the human skills needed to implement and maintain the technology. This also helps to set realistic expectations for data solutions that are consistent with the maturity of the required data sources, technologies, and capabilities.

5 - Building data solutions efficiently needs a scalable AI Factory and an agile development process

"As organizations seek to achieve tangible business results from their investments in data analytics and artificial intelligence, it's critical to adopt a focused approach that builds the right solutions and sets the right expectations. Through this approach, business leaders spearhead the development of data & AI solutions 'for the business by the business' - prioritizing the most impactful solutions, building quick POCs with data experts, scaling data solutions that work, and accepting 'failure' on those that don't. Having business teams lead the whole process ensures business buy-in and adoption by design."

Oussama Ahmad, Data Consulting Partner, Global Travel & Tourism Lead - ARTEFACT

Building and scaling data solutions for businesses requires a new operating model - an AI Factory - made up of feature teams led by business experts supported by data scientists, engineers, analysts and software engineers. This team structure ensures that data solutions are always built with a business objective in mind. Adopting an agile test-and-learn process that attempts to build a successful POC in a short time span is also essential to achieve faster time-to-build.

6 - Accepting that some data solutions will fail, and scaling and maintaining those that work

Not all data solutions will succeed; some will fail, due to technical or data limitations, despite careful planning and execution. It is crucial for organizations to recognize that failure is a natural part of the development process: it should not discourage them from pursuing future projects. Instead, companies should focus on industrializing successful data use cases, scaling them to full data domains, and optimizing their algorithms and data sources. This also includes ongoing monitoring and improvement of the use case to ensure that it continues to meet the needs of the business users.

7 - Sharing knowledge is necessary but not sufficient for wide data solution adoption

Providing data solution training and easy-to-use documentation for business users is necessary, but usually not sufficient for widespread adoption of data use cases. Widespread adoption of data solutions by business users is best achieved by having users lead the development process, integrating these solutions into the organization's learning curriculum, and including adoption and impact KPIs in business user scorecards. By aligning business user scorecards with the organization's

data strategy, organizations can create a culture of data-driven decision making and ensure that the adoption of data solutions leads to tangible business impact.

8 - Improving data solutions is continuous; prioritizing enhancements that matter is key

To achieve continuous enhancement of data solutions, it is vital to regularly collect feedback from business users, evaluate their needs and requirements, and make necessary adjustments to optimize these. The Scrum methodology provides an effective approach for gathering and implementing improvements in an iterative and incremental manner. Users of data solutions should log continuous feedback on the accuracy and usability of data solutions as well as required improvements to business processes. It's important to (1) implement improvements that increase the accuracy of the solution's output, (2) expand its features and functionality, and (3) improve its usability and user experience.

9 - Maintaining robust governance of data solutions ensures accurate results with minimal oversight

Maintaining high-quality data sources for data solutions is crucial for achieving automated, accurate results with minimal oversight. To achieve this, organizations should implement a robust data quality framework that enforces clear guidelines and standards for data collection and transformation. In addition, organizations should implement strong data security and privacy policies for secure and compliant data processing. This approach ensures that input data is accurate, current, and consistent, which reduces the risk of errors and improves the overall efficiency of the data processing workflow.

10 - Tracking the business impact of data solutions requires defining direct impact KPIs and assigning incremental business impact

Identifying the commercial or operational KPIs that are directly improved by a data solution is essential to measuring its business impact. Once these KPIs are identified, the next step is to develop a formula to measure the incremental impact of the data solution on each of these KPIs. This formula should take into account the baseline of these KPIs before (or without) the implementation of the data solution and compare it to the performance of these KPIs after (or with) the implementation of this solution, taking into account other factors that may have led to this increase. Once the incremental impact on each KPI has been calculated, it should be translated into financial terms, such as reduced costs or increased revenues. Finally, it's always recommended to use automated business impact measurement of data solutions to ensure unbiased and timely measurement of business impact.

"Data acceleration projects have been surging in the MENA region in recent years, as organizations embrace the power of data for business growth. While certain challenges persist, such as maintaining data quality, especially with legacy systems, organizations are actively seeking solutions to overcome these obstacles. Building the right data capabilities within business teams and the right operating model is the single most important way to ensure the successful implementation and adoption of data solutions and the realization of tangible business impact."

Karim Hayek, Data Consulting Senior Manager
ARTEFACT

From idea to implementation: becoming an AI factory

Alexandre Thion de la Chaume
Managing Partner
Data Factory - Industries
ARTEFACT

Formulating a coherent AI strategy, and deploying value-adding and efficient use cases is a struggle for many businesses. Alexandre Thion de la Chaume, Partner, Data Consulting at Artefact, explains how these processes can be streamlined through the AI Factory model.

Artificial Intelligence (AI) is seen as the major lever of competitive advantage. The data doesn't lie: there's been an almost 25% year-on-year increase in business use of AI, with 63% of executives agreeing it has led to revenue increases. The global pandemic has only put this into sharper focus. The businesses that thrive and survive will be those able to adopt the right AI solutions and deploy and scale them quickly and efficiently.

Yet, as with all game-changers, AI initiatives raise new challenges. Implementation comes with many questions – chief among them, how can you adopt the right data approach to deploy AI initiatives rapidly and efficiently, without failure and sustainably over the long term? The 'AI Factory' approach has been developed for precisely this reason.

The AI Factory is an organisational operating model – combining different talents, capabilities and processes in a systematised way – to deliver success in AI deployment and scalability. It has been effectively used by industry leaders like Carrefour and ENGIE to deliver transformative AI projects across their businesses. Yet setting up an effective AI Factory from scratch can be daunting. You need expert teams and a clear vision to make the process work.

Planning makes perfect

The vital first step is to define a vision and use cases for your AI Factory. This will be your data strategy. Use cases offering the highest business potential of transforming the company must be identified. Whether it's supply chain optimisation or compliance management, opportunities exist at all levels.

The company's AI vision should also be considered. It's important to have the ability to imagine how it could develop, to plan for it and to reach a clear-sighted idea of the future. From a preliminary overall view, draw a refined version applicable to data and AI.

The four pillars of the AI Factory

Once the company's data strategy and AI vision are defined, you should have a prioritised list of use cases to implement. But how can you start working on them? An effective AI Factory implementation is founded on four distinct pillars:

Next, concrete business opportunities must be assessed, through the identification and sorting of use cases. This is done by assessing business impact and implementation complexity. A focus on mindsets is important throughout, to manage change on a large scale and involve everyone, from company leadership to front-line team members.

ONE SINGLE GOVERNANCE

To be efficient, governance must be high-level, dedicated and tailored. A highest-instance AI Factory Board – comprising key C-suite data leaders – is extremely important in providing overall sponsorship and direction as it shares the AI vision and aligns teams and the roadmap with it. At the programme management level, an AI Factory Director role should be established, involving business, operations, legal, security and IT data experts. Their role should be to review, arbitrate and validate progress.

Finally, at the operational level, there need to be agile teams. Feature Teams are responsible for the delivery of use cases with AI products. They're close-knit units working collaboratively to ensure permanent information flow and transparency. Most importantly, they should be multidisciplinary, combining skills and expertise from across the business. They are achievement-oriented, each one created with a single objective: to deliver one use case measured by a unique goal.

ORGANISED, DIVERSE AND EXPERT TEAMS

To drive efficiencies, structured organisations should gather business, data, software and digital tech skills in

hybrid teams based on agile methods. Agility ensures a flexible and adaptive way of working and avoids issues linked to a silo approach, such as isolated departments within the same structure or overly rigid procedures. This requires a good blend of business and technical profiles, to ensure that what is developed on the technical side always has a useful purpose that addresses business needs.

Scalability is an important overall characteristic of a team's makeup. The idea is that its structure can be easily duplicated, like Lego bricks. With a fully scalable model, more teams can be added to address additional use cases.

ADVANCED AI TECHNOLOGIES

Of course, effective AI deployment needs a foundation of AI-enabling technologies. An AI Factory uses a combination of open-source, proprietary and cloud solutions. They should be standardised across the whole data pipeline – from ingestion to visualisation – from beginning to end, according to best practices.

SYSTEMATIC & PROVEN METHODOLOGIES

Systematisation is needed to make sure a series of steps are always taken in a specific order, each with its own defined objective. The benefits are twofold. First, this gives an overall structure of common references throughout, creating a backbone that guarantees consistency. Second, this makes methodologies replicable and scalable, considerably accelerating the deployment of the industrialisation phase.

MLOPS: KEEPING THE FACTORY RUNNING

Alongside a set use case methodology, MLOps (Machine Learning Operations) practices must be deployed to close the gap between the concept phase and production. Inspired by the DevOps process, this should combine software development and IT operations to shorten the development life cycle.

The purpose of MLOps is to tackle challenges that traditional coded systems do not have. The first challenge is collaboration between teams: different units are often siloed and own different parts of the process. This stifles the unity needed to go into production.

The second is pipeline management, as ML pipelines are more complex than traditional ones. They have specific characteristics, including bricks that must be tested and monitored throughout production.

The final obstacle is that ML models usually need several iterations – when put into production in a manual, ad-hoc way, they become rigid and difficult to update.

Instead, an MLOps approach should embed all ML assets in a Continuous Integration and Continuous Delivery pipeline (CICD) to secure fast and seamless rollouts. All data, features and models should be tested before every new release to prevent quality or performance drift. All stakeholders should work on the same canvas and apply software engineering best practices to data science projects – versioning, deployment environments, testing.

Ultimately, MLOps is the discipline of consistently managing ML projects in a way that's unified with all other production elements. It secures an efficient technical delivery from use case early stage (first models) to use case industrialisation.

A FRAMEWORK FOR SUCCESS

AI holds tremendous promise, but also great risk for organisations unable to deploy it properly. The real benefit of the AI Factory model is that it establishes a core framework for swift and successful implementation. Processes, teams and tools are transferable and repeatable by nature, meaning a company can remain agile in pursuing its AI vision. Once the process is established and supported by MLOps, a business has what it needs to become an AI powerhouse.

CASE STUDY

CARREFOUR
Using AI to drive value in store based on the AI Factory's operating model of Artefact

CHALLENGES

AI as a corporate strategy.

AI offers incredible opportunities in the retail space. Global retailer Carrefour is going through a digital transformation and has partnered with Google and Artefact to leverage the power of AI and capture value in several departments: assortment, pricing, supply chain, store operations, ecommerce, and marketing.

"We aim to build Artificial Intelligence and Machine Learning solutions to better serve our customers and employees"

Elina Ashkinazi-Ildis — *Director, Carrefour-Google Data Lab*

Carrefour's ambition is to sift through its vast trove of data (4 billions annual transactions, 1 million daily visits to digital platforms) to identify unaddressed issues, define use cases, scale AI solutions, spread the adoption of AI within the company and conduct training and upskilling.

"We are really trying to inject innovation, agility, extra-collaboration"

Amélie Oudéa-Castéra — *Head of E-Commerce, Data and Digital, Carrefour*

Carrefour chose to set up a multidisciplinary hub of internal and external data experts.

40%
additional revenue

AI Factory by Artefact, a robust framework that turns AI technology into valuable AI projects and solid competitive advantages.

Artefact devised an operating model through an agile methodology composed of several steps: Structure, Discovery, Minimum Viable Model (MVM), Prototype, Scale and Optimization.

"The key is to carefully select the right use cases that bring value to the business."

Vincent Luciani — *Co-CEO, Artefact*

Carrefour selected a dedicated AI lead and worked with its partners Google and Artefact to establish guidelines. Artefact also assembled Feature Teams, each working on a single unique key performance indicator. They are made of a business owner, an AI product owner, a data engineer and a data scientist.

"When a challenge is huge, our ambition is to break it into many sub-problems that we will solve one after the other"

Vincent Luciani — *Co-CEO, Artefact*

Use cases were developed in several departments, such as assortment optimization, dynamic pricing, relevant promotions, sales prediction, inventory management, out of stock prevention, fraud optimization, customization of marketing, churn reduction, and algorithm product recommendation.

"Our rule is the golden KPI: to define a numbered objective that is very concrete. For example, regarding supply chain, our KPI is forecast accuracy."

Fabrice Henry — *Managing Partner Data & Consulting, Artefact*

Relying on a gradual approach to gain speed and scale while enabling innovation is key to turn experimentation into innovation.

Making data-driven decisions and unlocking value at scale.

Artefact's initial experiments proved successful and are being scaled through the organization and deployed across different product categories, store formats, and countries.

Carrefour developed an assortment recommendation tool that helped the chain support a more personalized selection at the store level, giving store directors the autonomy to influence inventory needs.

Some stores saw up to 40% additional revenue on some single items.

"Carrefour needed to ensure it had the right products, in front of the right shoppers, at the right store location"

Stephane Spinella
Retail Director, Google Cloud

The models made to optimize operations with precision across the supply chain were able to detect stockouts in just an hour when it used to take two days and to accurately predict the volume of curb pickup sales down to the half-hour and allow store managers to staff their operations teams accordingly.

"We are redefining the ways of shopping, developing a truly omnichannel value proposition for our customers … Across the continent, there is value creation in our core business"

Amélie Oudéa-Castéra
Head of E-Commerce, Data and Digital, Carrefour

AI factories are a combination of talented individuals, methods, and technologies in the service of brands looking for scalable operational efficiencies and business successes.

CASE STUDY

HEINEKEN
Using the Data Factory methodology as a Revenue Generation Center

CHALLENGES

HEINEKEN Brazil had an ambitious challenge: add business value through the use of data and advanced analytics. So, Artefact joined us to accelerate, and make this happen.

SOLUTION

Rafael Melo – *partner* **ARTEFACT**

We implement our Data Factory methodology, which are hybrid teams composed of business experts, data scientists and engineers, to deliver a product that is quickly actionable.

The teams are a mix of people from Artefact and HEINEKEN, to help in the Data Driven acculturation which was one of HEINEKEN objectives.

Artefact delivers data products from start to finish, from the business problem prioritization phase and its solution, through data mapping, collection, and exploration, creation of machine learning model and a final product to activate this model.

Finally, we test the solution and industrialize this product for larger scopes. For this, we always rely on agile principles: We start with a reduced scope, to quickly show business value to stakeholders, and developing the solution incrementally.

In this partnership with HEINEKEN we created data products in practically the entire value chain: Such as finance, HR; production; distribution & logistics; marketing and trade, as well as sales and e-commerce.

![Heineken logo]

Daniel Guimarães
Logistics & Planning Manager
HEINEKEN BRAZIL

We had a challenge in the area of planning and logistics related to allocating products in distribution centers and making short-term decisions. The challenge was both extracting the information and creating the intelligence to generate the insight needed daily.

In this way, we developed a stockout prediction model, which consists of the complete automation of data, modeling and creation of a dashboard which generates the necessary insights for our decision making.

We started for a few products and a few distribution centers, but quickly saw the value of the solution and scaled to the rest. Today, this model is one of the main decision-making tools in the area.

Camila Moreno
Data Scientist - **ARTEFACT**

HEINEKEN invests heavily in the automation of its factories, and one of the projects with Artefact was to use sensor data, like temperature, pressure and volume, to create a machine learning model that makes adjustments to production, even during the process, ensuring quality metrics, such as, the color of beer.

The interesting thing about this type of project is that the automation and financial gains can be easily scaled to other breweries

RESULTS

Fábio Criniti
Data & Analytics Director
HEINEKEN BRAZIL

The biggest benefit of this partnership with Artefact is the speed at which we are able to deliver value to the business, and build a revenue generation center for HEINEKEN.

Hybrid teams are able to very well connect the problem with a data solution. For us this is very important, as we were able to prove value and consequently invest more in innovative projects like these.

23

Data Governance, a prerequisite for AI project success

Justine Nerce
Managing Partner
ARTEFACT

Data and its applications are being increasingly integrated into business activities. They're at the heart of the search to improve productivity and overall efficiency. Through specific processes and an adapted organizational structure, data governance enables companies to organize data, enhance its quality and meet the ethical and regulatory challenges of data processing. An interview by the Hub Institute with Justine Nerce, Partner at Artefact.

What are the challenges of data governance today?

The amount of data and the number of use cases around data is constantly increasing. First, companies have to deal with the challenge of getting the most possible value out of their data and democratizing it. Good quality, well documented data should allow it to be accessible to the end user.

All of this applies within the framework of ethics and data protection. Data governance is becoming essential to ensure compliance with certain privacy laws. In Europe, the General Data Protection Regulation (GDPR) is in effect and is tending towards becoming the global standard.

This means that the organization must first be able to demonstrate that it knows what data is flowing through its infrastructure. It must be fully transparent about what data it is collecting from its users and be able to delete all data linked to any individual immediately.

Second, migration to the cloud is essential. Three main uses are emerging: Business Intelligence, Artificial Intelligence, and data exploration. By structuring data based on data governance, businesses will be able to offer these three types of uses as a service.

These data products constitute a common, cross-disciplinary good, which requires a dedicated team. These products must be of high quality, but also visible and usable by all. The challenge for companies is data democratization. These data products must also be secure and protected to comply with various regulatory and ethical issues.

How does Artefact support companies in implementing data governance?

At Artefact, we act as a consulting firm. We support all our clients throughout
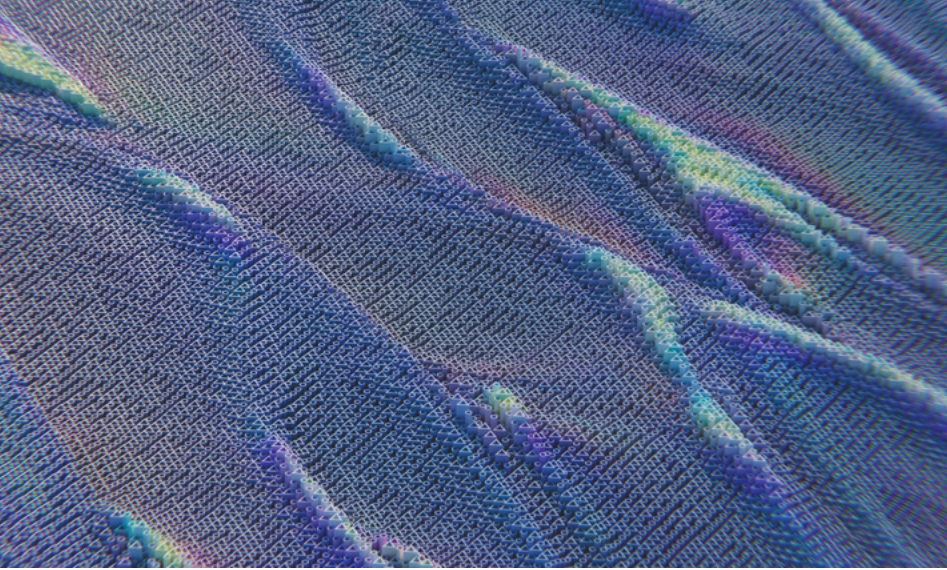
data governance implementation, from strategy to deployment. First, we perform an audit to see where they stand, then define a roadmap to identify areas to work on. Finally, we build a data asset structure into data products and help them choose the technical tools they need.

In our consulting approach, we insist on the importance of data as a vector of value for the company, then we work on deployment, quality tool selection and documentation of governance to give substance to the strategy and make it feasible.

We've also set up our own Artefact School of Data, which lets us train data stewards and data owners, essential roles in the implementation of data governance for businesses. Along with this professional training, we also intervene directly in companies to acculturate them to the need for advanced and supported data governance in order to succeed in their AI projects.

What is unique about Artefact's global vision?

Our strength is that we propose a global data governance model, focusing on end-use cases first. We position data governance as an "asset" of this transformation. We're able to transcribe use cases into tangible value and be part of a global transformation program.

We also have multidisciplinary experts. There are about 20 of us in France who specialize in data governance, with profiles from different backgrounds: data product owners who model products, data stewards who document and improve quality, but also data engineers and data analysts.

We also have an ecosystem of technology partners with whom we collaborate in an agnostic way. We're proficient in all the new tools that appear on the market. We have both technical and strategic DNA, and are able to link all of these subjects together to treat them in a holistic and comprehensive way and deploy them to many clients.

Have you got a concrete example of support that you've provided?

We assisted one of our major clients with very extensive data assets in their data transformation. The project concerned a redesign of their data governance. When we arrived in mid-2017, we saw that their governance had been approached from a too-technical and not sufficiently "business" perspective. This resulted in a lack of adoption of the necessary tools. To correct this, we linked their governance to their strategic use cases. To do so, we documented the use cases, democratized their access, and improved data quality to ensure good results. The first pilots were a success! We then faced the challenge of scaling up.

In 2020, we assisted this same company in launching a program to accelerate Artificial Intelligence programs and migration to the Google Cloud Platform (GCP). Governance had been positioned as a strategic asset of their transformation and this launch was performed in two stages:

- **Structuring their "Data Governance Office"** and setting up an operating model with data stewards and data custodians, etc.

- **Structuring their data assets into a large "business domain",** with the choice of tools to operate, etc.

We're now entering a third phase of industrialization and extension of this AI program. As part of the migration to the cloud, we're analyzing how we can structure, rationalize and pool our data assets. At the moment, we've moved on to the second stage, which consists of structuring our data assets according to these major business families. Next, we're going to start thinking about the development of tomorrow's data products, which will serve different categories of use cases.

What can we expect in the future, once everyone has implemented their data governance?

The availability of data will allow the implementation of even more use cases, particularly in the area of Artificial Intelligence. This will accelerate value creation within organizations. It will also allow us to support all the issues surrounding data democratization and decentralization, especially in terms of bringing data closer to the business. Artefact's mission is to create this bridge between data and business, and we carry it out on a daily basis with our clients. If the data is well structured and clean, if the products are available, and if we have the push-button tools to manipulate them, theoretically in five years, everyone will be able to use data in their daily work!

25

ARTEFACT

The vital role data governance plays in achieving sustainability goals

In this article, we will present how to define sustainability goals and how to include them in data governance strategies.

Manuela Mesa
Director
ARTEFACT

Natacha Zouein
Senior Data Consultant
ARTEFACT

Pauline Billerot
Data Consultant
ARTEFACT

Sustainability is a key focus for today's organisations, and with consumers' purchase decisions increasingly based on 'green' credentials, it can be a critical element in remaining competitive. Businesses are starting to improve their sustainable practices by addressing the products and services they provide, the processes they use, the waste they generate as a by-product, and the supply chain that facilitates their operations. But while 90% of executives believe that sustainability is essential, only 60% of organisations have sustainability strategies in place.

In the data-driven world, companies have a wide range of effective tools at their disposal that can turn data into value to accelerate the implementation of a sustainability strategy. They can collect and examine data on a wide range of sustainability-related issues – from energy use to carbon emissions – to reveal key insights that drive initiatives. In addition to enabling green capabilities, analysis shows that, on average, every dollar invested in data results in $32 in economic benefit. In other words, ensuring that data is accurate and reliable is essential for organisations.

Focusing on creating, maintaining and securing high-quality data is key, but equally important is ensuring that this data is accessible for the analysis that enables data-driven decision making. Consequently, it is crucial to adopt strong data governance – a set of processes and policies that can be implemented to ensure data is reliable and trustworthy.

Defining sustainability goals: a challenging road for companies

In 2015, the United Nations presented its 17 Sustainable Development Goals (SDGs) as the blueprint to achieve a better and more sustainable future for all; it expects companies to have established sustainability strategies and implementations in response to them by 2030.

SDG-oriented business models have the potential to create significant market opportunities. In 2019, McKinsey estimated that global sustainable investment had exceeded $30 trillion, a tenfold increase since 2004.

Failure to define SDG goals and apply them to their business model puts companies at risk in three core areas:

1. **Financial Companies can face enormous costs due to environmental risks** that affect their supply chain. For example, Unilever estimated an annual loss of €300 million due to climate change endangering agricultural productivity; the company is currently working on a pilot project (using SAP's GreenToken supply chain transparency technology) to further increase traceability and transparency of its global palm oil supply chain.

2. **Legal, compliance and risk management**: Different countries have different regulations, which may lead to confusion and even risk. In the UK, various laws and frameworks require organisations to be transparent in areas such as diversity, equal pay, carbon emissions and modern slavery. The Competition and Markets Authority (CMA) guidelines, released in January 2021, helps businesses understand the rules that apply to their operations and how to achieve sustainability goals without breaching competition law.

3. **Customer trust**: Today's consumers are actively choosing brands based on their ethical behaviour and their initiatives linked to sustainability and climate change – although 48% of UK adults say they do not trust the information companies provide around sustainable products, indicating a risk factor for the relevant companies. In addition, 36% of people in the UK believe further regulation to make companies improve sustainable lifestyle choices for consumers should be introduced.

However, firms are facing challenges in defining and implementing their sustainability goals. One recurring obstacle is ensuring that the adoption of a sustainable strategy will not impact their profitability. Businesses need a quick return on investment, and a company must be profitable to be sustainable. At the same time, the work of measuring ESG scores may prevent some executives from fully investing in sustainability initiatives, as 63% of CEOs struggle to measure ESG across the value chain, representing a barrier to sustainability in their industry.

Companies need to identify relevant KPIs to create valuable sustainable insights. By measuring these KPIs, companies will have opportunities to achieve their ESG goals, such as carbon footprint reduction, energy consumption, waste and pollution tracking (i.e. within the supply chain), and social impact. But to accurately measure these KPIs, organisations must be able to rely on trusted data to create tangible results, and accessibility to the relevant data can be hard to gain. For example, if greenhouse gas (GHG) emissions reduction is identified as a KPI, sustainability teams will need to access hard-to-get financial data, such as travel mileage, and combine it with human resources data to calculate the GHG emissions of individual employees.

Sustainability strategies and goals are crucial for companies and if reliable data isn't available and accessible, their societal, environmental and legal requirements won't be met. Companies cannot implement sustainability strategies without data

governance that offers transparent and valuable data for better data-driven decisions.

Data governance: what it is and why every company needs it

Data governance is the approach companies take to set standards and policies on how data is ingested, processed and used in a way that makes it secure, available, accurate and usable. It includes aligning the people, processes and technologies needed to support those standards. Putting a data governance policy in place provides businesses with a formal strategy with which to access, monitor and use data to support employees and business units. It highlights data's role as a valuable asset that is essential to respond to strategic needs and enable data-driven decision-making, resulting in the following benefits:

•**BETTER DATA QUALITY:** Accurate and reliable data provides companies with a tangible business asset. Using clean data brings business processes across the company into line with each other; this compatibility results in reliable performance measurement and dependable KPIs.

•**COST AND TIME SAVINGS:** By applying greater data management discipline through better visibility and standardisation of processes, companies can redeploy 35% of their data spend. Moreover, reliable and accessible data saves time by reducing manual tasks.

•**BREAKING DOWN SILOS:** Avoiding data duplication, outdated or incorrect information and silos (collections of data that are isolated across different business groups), reduces storage costs and, most importantly, increases operational efficiency. (Artefact's experience shows that data scientists can spend more than 30% of their time on understanding and accessing data.)

•**COMPLIANCE:** A data governance framework provides companies with data security and enables them to meet compliance regulations (such as GDPR) and stay on top of their legal obligations; data governance is designed to help companies operate more efficiently.

Implementing a data governance strategy

There is no such thing as a one-size-fits-all approach to data governance. Each strategy is unique to the organisation it serves and requires a different solution. To best define the optimum data governance for the company in question, a framework should be followed; based on its long experience with clients, Artefact proposes the following approach:

1. **Vision and business requirements:** Define the company's business priorities and objectives, as well as its vision for data strategy in the short, medium and long term.

2. **Data infrastructure:** Identify where the organisation's data currently sits, whether the infrastructure is designed to facilitate operations, and whether data is constantly updated.

3. **Current data governance:** Establish whether the company's data is supervised, stored securely and easily accessible, and if it is uniform across the organisation.

4. **Data applications:** Define what is required to achieve the organisation's vision. Solid application development processes are essential (development, prototyping, industrialisation), and all applications need to be linked to business objectives and add value across the company.

5. **Monitoring and evaluation:** Ensure continuous checking of the data objectives through clear KPIs and targets.

6. **People and processes:** Put the right processes in place to implement the data strategy.

7. **Tools and capabilities:** Ensure that the right tools are used, and up to date, to facilitate data processes and enable the required changes.

Organisations need to set up a data governance programme, which should involve structuring data governance assets (definition of the operating model, tooling and roadmap of the
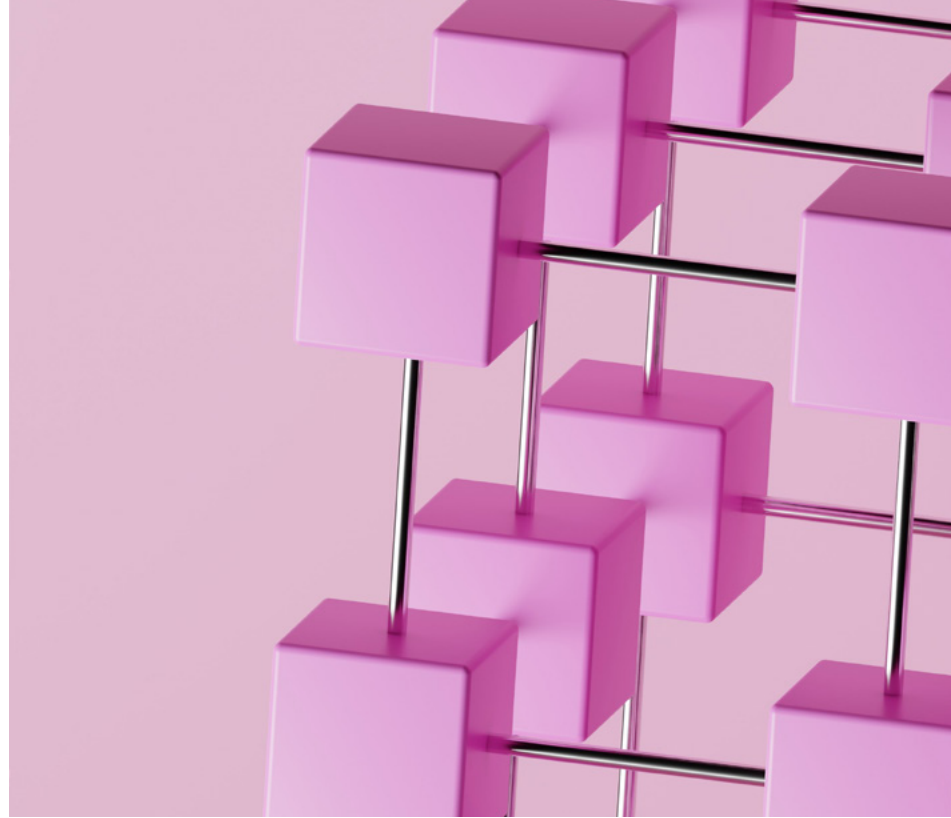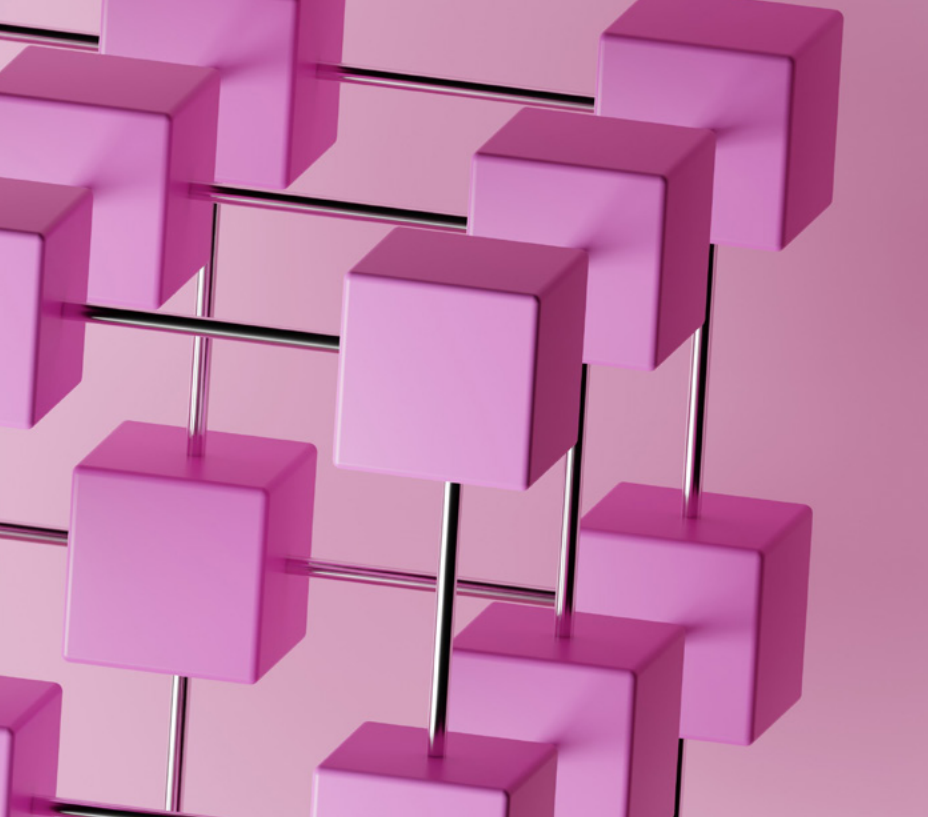
data governance initiative) and the deployment of data governance within each domain (data quality, standardisation and accessibility).

Once companies have set up solid data governance and defined their sustainability goals, the next step is to identify how to leverage the first to achieve the second.

Using data governance to achieve sustainability goals

Advanced technologies can use data to uncover deep insights, opening a world of innovative ways to support sustainable practices across the enterprise. Artefact has worked with several organisations to build data/AI products and strategies, all of which are based on strong data governance, that integrate with business processes to tackle energy and environmental issues.

ENERGY CONSUMPTION REDUCTION:

Artefact supported a leading European telecommunications provider to address an environmental initiative to decommission copper across its network by 2030. The sustainability goals were to deploy a copper shutdown programme,

minimise the energy consumption of copper while it was still being used, and quantify the risk levels of the project using AI. A tool was created to prioritise geographical areas for the work to be carried out and optimise costs for the initiative. As a result, the company estimated that it could save €1.15M in costs, 1.65 GWh in energy consumption, and 111 tonnes of eCO2 on average per year between 2025 and 2030.

INCLUDING SUSTAINABILITY USE CASES IN DATA STRATEGIES:

Companies are incorporating sustainability objectives within their data strategies, such as gathering energy and utility data from their facilities to work towards carbon emission reduction. For example, Artefact helped two leading property management companies in the United Arab Emirates (UAE) to understand their data and define clear sustainability objectives to achieve their goals in this area. The projects looked specifically at facility, utility and energy waste. One company partnered with third party providers to implement AI and smart technology across shopping malls to track air conditioning and electricity consumption and pinpoint

how to reduce usage and save money. Additionally, dashboards to monitor and manage utility and energy waste were created, helping one of the companies with its main goal of identifying where solar panels could be installed to save costs in shopping malls and residential properties. Identifying sustainable objectives within their data strategies provided both companies with significant benefits; one forecasted 6% revenue growth over the next eight years.

WASTE REDUCTION

French retail giant Carrefour had issues with stock availability/ shortages and shrinkage in its bakery department due to a limited ability to predict consumption on any given day. Artefact provided Carrefour with a forecasting model that generates daily predictions for each store and product line. Integrated into the current tool as enriched information, it provides each store manager with reliable predictions so that daily production of baked goods can be adjusted accordingly. The project let Carrefour reduce waste on fresh bakery by 12%.

Additionally, Artefact developed a tool allowing Carrefour to measure and model the carbon emissions of its e-commerce sales in 2021, from click to delivery. It was able to measure 100% of emissions in four weeks and design a dashboard for simulation and monitoring.

SUSTAINABILITY STRATEGIES NEED STRONG DATA GOVERNANCE

Data is a vital lever for achieving sustainability goals, but it needs proactive management if organisations are to accurately measure their impact in this area.

Structured data governance should therefore be an integral part of any sustainability strategy; once in place, companies will be able to lay the foundations for transparent and accurate decision making and derive real business value

Group
Pierre & Vacances
CenterParcs

PIERRE & VACANCES CENTER PARCS

How data governance and data quality can boost digital marketing and activation performance

CHALLENGES

Improving data quality across the entire enterprise.

PVCP needed to be able to exploit all the data available in their ecosystem. To do so, they had to be able to:

- **Improve** their data quality to make it more reliable
- **Join and structure** data so it's clean and can be shared
- **Help businesses** democratise data use and exploit this shared data for personalisation to build enhanced customer relationships.

Artefact knew that PVCP was going to need help to recoup revenues lost due to the COVID-19 crisis: a 3% drop in worldwide tourism industry growth was forecast for 2020, while there was a total of 40B€ estimated lost revenue during the lockdown in the tourism sector in France over the last year. But some regions are expecting to see modest rises in tourism in the coming year.

"An interesting point to highlight about this project is the use of this unique period we're living through – it's an ideal time to invest in data quality and data governance – because these are prerequisites and fundamentals, we all need if we are to use this period intelligently and anticipate the future."

Fabien Cros,
Data Consulting Director at Artefact, added.

To meet the objectives of PVPC, it was necessary to:

- **Prioritise Data Quality and Governance** above all other subjects to work efficiently and reliably in order to have an immediate impact on the business

- **Form a committed SWAT team** composed of experts ready to work hand-in-hand with PVCP on complex subjects in order to correct current data quality issues – but also to prevent new ones

- **Create a new Data Steward role** with a network of SPOCs (Single Points of Contact) to reduce quality problems and produce high-performance analytics available to all departments.

"The objectives of the data quality project with Artefact were simply to have a better cohesion in the quality of the data and to set up processes to help us be more efficient in the way we deal with the different subjects,"

clarifies Julien SOULARD,
PVCP's newly-appointed Tracking & Data Collection Specialist.

SOLUTION

A SWAT team to roll out a Data Governance project.

PVCP and Artefact worked closely for several months on these complex data science projects.

To lay the essential groundwork, each company needed to be able to work in a SWAT team configuration. SWAT is an American military term meaning Special Weapons And Tactics, but in the business world, a SWAT team is composed of experts in various fields that come together to rapidly and efficiently validate new business ideas within an organisation.

The PVCP-Artefact SWAT team:

- **Prioritised** and ultimately delivered value more quickly

- **Mobilised** different skills to optimise global knowledge

- **Targeted** only one value chain at a time for optimal efficiency

- **Delivered** with a stronger focus on results

The PVCP-Artefact SWAT team was dedicated to generating more high-quality sales leads, optimising the media budget, and improving processes and workflow.

With the team in place, a major Data Quality Project was begun, composed of Data Governance, Tool Creation and Utilisation, and Monitoring, with speed and efficiency as driving factors.

To kick-start the Data Governance facet, a PVCP – Artefact Data Steward pairing was proposed, while the Data Steward and SPOCs were trained on using data quality tools and methodologies.

The Data Steward is responsible, among other things, defining data quality standards, monitoring KPIs, and holding a weekly update to monitor and investigate all current data issues with the data quality SPOCs.

To address these issues, several tools were created, most importantly the "golden source" – a data quality dashboard which tracks every KPI and signals every anomaly. There's also a ticketing tool set up which enhances communication when it comes to the way queries are posed.

"Today the teams are 100% autonomous in all their roles."

Clara Mendes Sampaio,
Artefact's Senior Data Consultant who paired with Julien in his new role as Data Steward, affirmed that

ARTEFACT

RESULT

Time & money savings, greater accuracy, happier employees.

The project rapidly showed appreciable benefits:

- **+28% improvement** in algorithm accuracy for improved decision-making

- **44% increase** in time saved on data comprehension, location and accessibility

- **+30 employee NPS points** added after implementation of the data governance frame (shift to higher added-value activities for people)

The project also improved the data quality on PVCP's analysis and piloting tools, thanks to the implementation of different processes, including weekly appointments, Mantis tickets and a dashboard for monitoring various tools.

"Before, our data was a bit like Swiss cheese: full of holes. [...] We'll continue to work with PVCP on new projects over the coming year: Google Analytics 4, GTM Server Side, that sort of thing. But now, we have a much more sound basis that will allow us to approach data problems more efficiently."
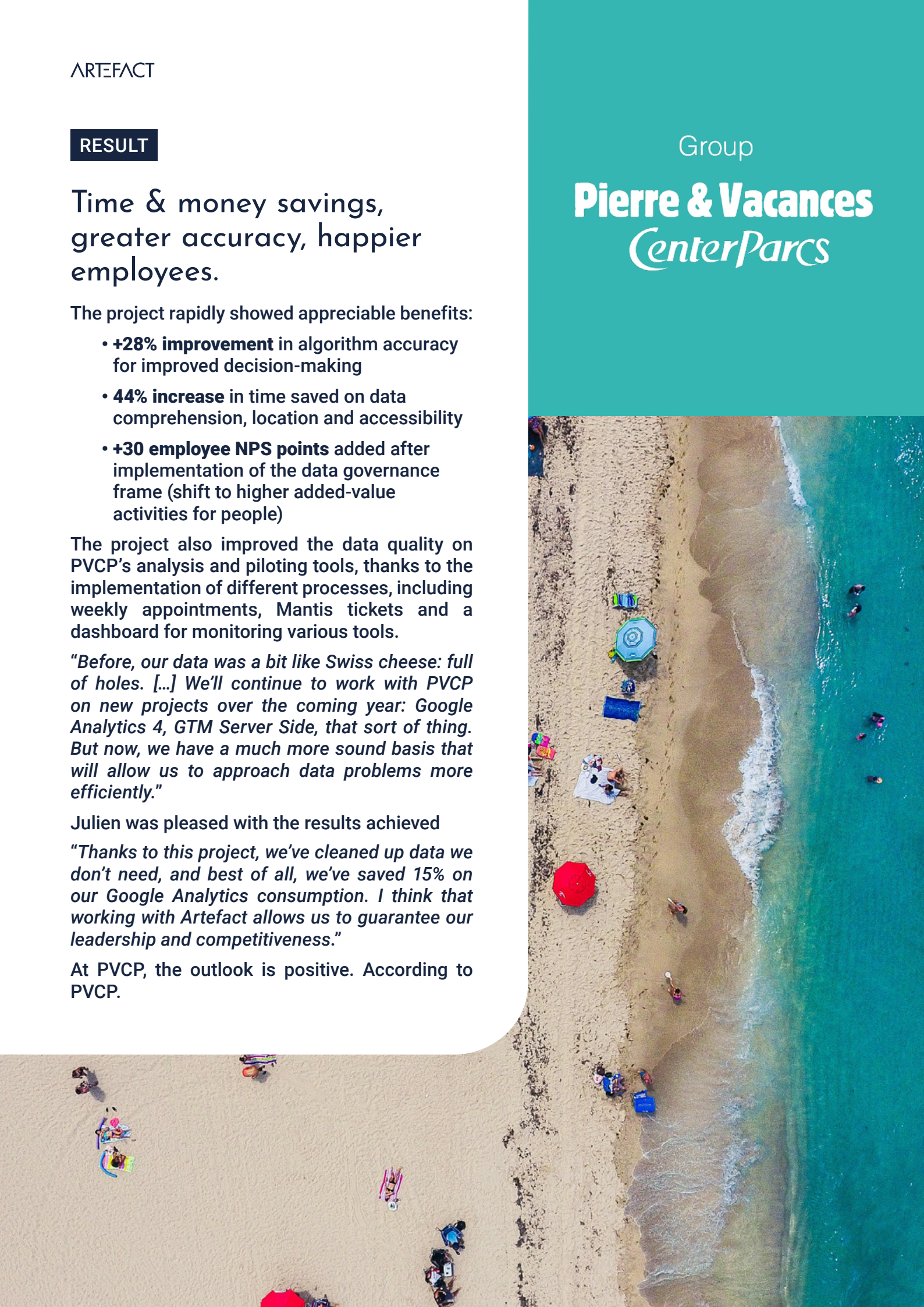
Julien was pleased with the results achieved

"Thanks to this project, we've cleaned up data we don't need, and best of all, we've saved 15% on our Google Analytics consumption. I think that working with Artefact allows us to guarantee our leadership and competitiveness."

At PVCP, the outlook is positive. According to PVCP.

Group

Pierre & Vacances
CenterParcs

Data Mesh: Principles, promises and realities of a decentralized data management model

Amine Mokhtari
Customer Engineer, Data Analytics specialist
GOOGLE CLOUD

Justine Nerce
Managing Partner
ARTEFACT

Killian Gaumont
*Senior Consulting Manager*
ARTEFACT

On 27 September at the Big Data & AI Paris 2022 Conference, Justine Nerce, Data Consulting Partner at Artefact and Killian Gaumont, Data Consulting Manager at Artefact, along with Amine Mokhtari, Data Analytics Specialist at Google Cloud, conducted a Data Mesh Workshop. Data mesh is one of the hottest topics in the data industry today. But what is it? What are its business benefits? And above all, how can companies successfully deploy it across their organizations?

Data mesh is a new organizational and technological model for decentralized data management. A distributed architecture approach for managing analytical data, it allows users to easily access and query data where it resides, without first transporting it to a data lake or warehouse. Data mesh is based on four core principles:

- **Domain-oriented data ownership,**

- **Data as a product,**

- **Federated data governance,**

- **Self-serve data as a platform.**

The workshop was divided into three parts:

1. **Business value:** Why adopt a product/mesh approach? How does it serve the company's business objectives?

2. **Deployment approach:** How to achieve success? What steps should be taken and what organizational model should be used?
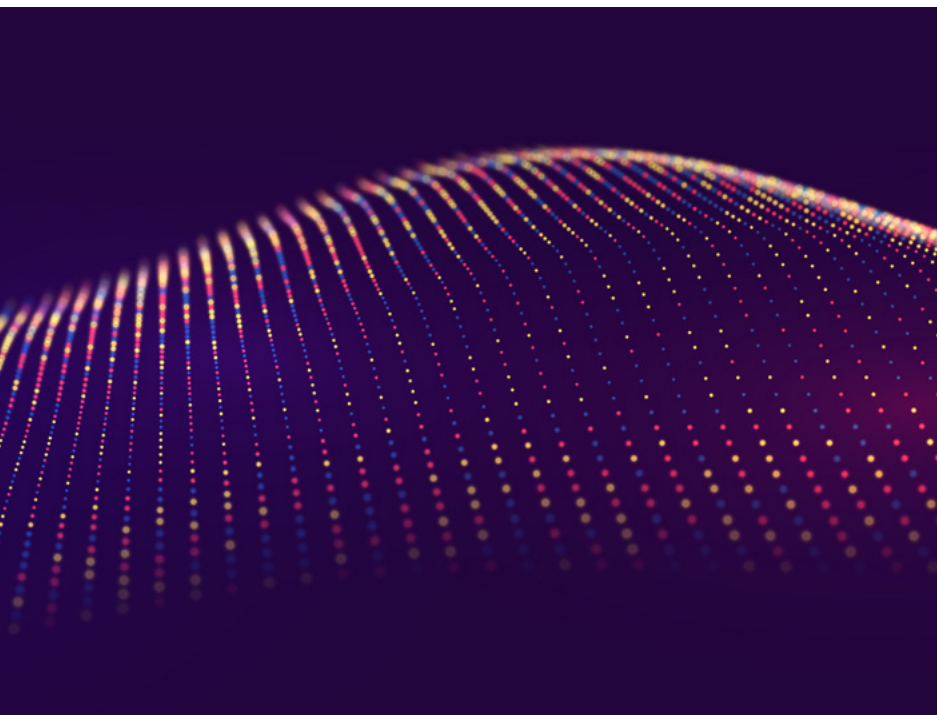
3. Technology stack: Why choose Google as a technology solution?

To kick off the Business value discussion, Justine Nerce explained: "One of the best reasons for adopting a product/mesh approach is that it eliminates two vicious circles. The first is 'reinventing the wheel' each time a new use for data emerges: a new team is formed that creates its own data pipeline to serve its specific needs. The result? Zero shareability, zero reusability for the technologies chosen. The second is 'building a monolith' when a new use for data ends up in the backlog of a central data team, then gets handed off to non-data specialist teams that carry out massive data collection, generic transformation and use case development, with the risk of not responding to user needs."
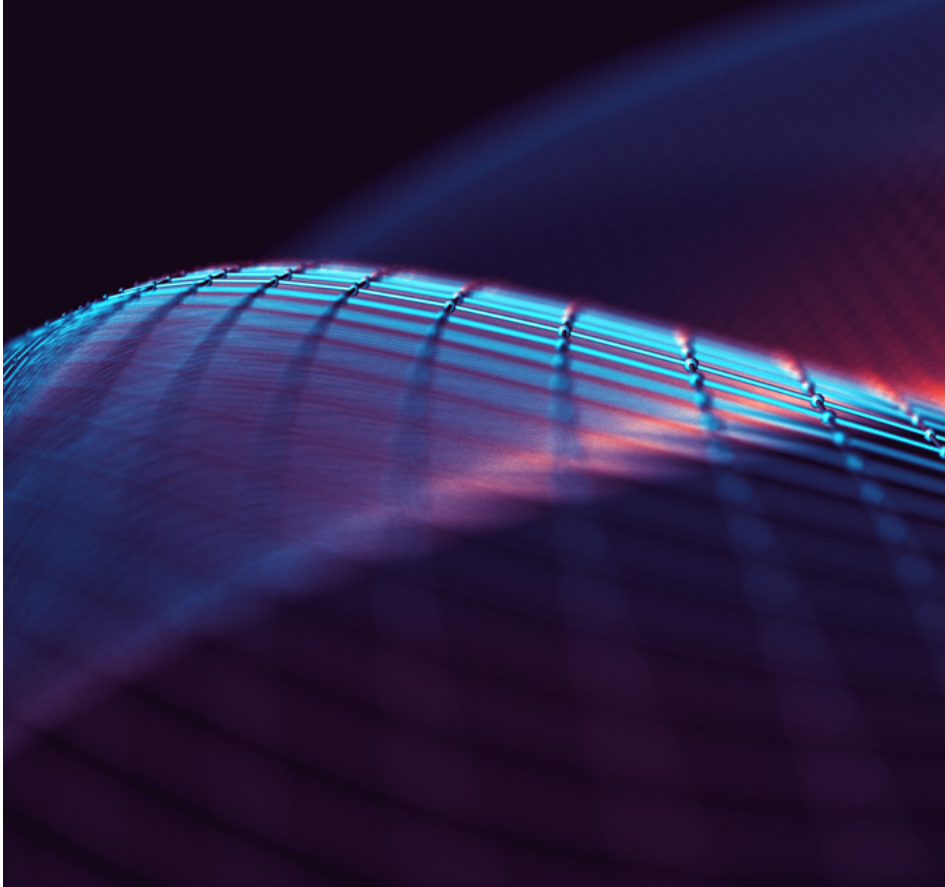
But with a product approach, the vicious circle becomes a virtuous one. When a new use for data emerges, instead of building something new, data mesh seeks out what already exists and can be reused. It identifies domains already in charge of handling given subjects and looks for existing data products that can accelerate the creation and development of new needs, either as they are or in iterative processes to create new, customized products. And all of these products can be published in the company catalog.

How data products create business value

Data products have existed in enterprises for a long time, but in data mesh, the uses and qualifications of data are essentially different, explains Killian Gaumont: "Today's data product is a combination of data made available to the business for business use and specific features that facilitate the use and the reusability of data".

To be included in data mesh, a data product must be:

"Today's data product is a combination of data made available to the business for business use and specific features that facilitate the use and the reusability of data".

1. **Governed by a team of dedicated owners;**

2. **End-user oriented and widely adopted;**

3. **Of quality throughout its life cycle;**

4. **Reusable as is or for building other products;**

5. **Accessible to all users;**

6. **Standardized so that everyone speaks the same language.**

At Artefact, data products are categorized into three different product families. "There are raw products such as databases used for business processes – which are data products nonetheless", assures Killian. "Next, there are data products enriched with customized algorithms or product recommendations, such as Interaction 360°. At the top are finished products aligned with use, such as dashboards. These are consumer-line products, designed to create value by linking product development to business strategy."

Prerequisite #3: Continuously def1ne data domains and scale up as soon as the model has proven its value

1

How to define data domains and associated responsibilities?

Start from data usage and existing systems

1. Map systems and processes and the people attached to these systems

2. Map business uses around data (BI solutions, applications) and main users

3. Establish first mapping and a common vocabulary

2

How to measure success on a first domain?

Measure success around 2 key axes

▼ ▼

Business use

Increase in average number of users of data products made available

Time to insight

Time to deploy a new data product or access information

3

When and how to scate up the model?

The value of data mesh organization lies by definition in the number of usersand in scaling up

Several signs that scaling up is possible

1. If the first data products are widely distributed and reusable

2. When a data domain needs a new data product

3. When a domain has the resources to build its own feature teams product

Deploying data mesh across the enterprise

Artefact's approach to data mesh deployment starts small, by prioritizing the business's use cases and pain points. All the domains and data products needed for each prioritized business use case (from raw data to finished products) are then identified. A future team is assembled to develop the first products and set standards. Then, related products to be built in the future can be identified.

There are three prerequisites for data mesh deployment. The first: breaking down silos.

"If data mesh is to be a success, we must move towards an organizational model that breaks down the silos between IT, data and business to have platform teams composed of cross domain and cross product teams, across all entities", says

Killian. "It won't happen overnight, obviously. But we've already begun breaking down silos by integrating business teams into IT data teams so that product teams developing data products can work more efficiently."

The second prerequisite is the Data Product Owner, who plays a key role in coordinating data mesh implementation. The data product owner has three missions: to design, build and promote data products. The first two missions are self-explanatory; the third is equally important, as the strength of a data product lies in the fact that it is adopted and used by the business. "The data product owner is responsible for ensuring that the data product is documented, understandable and accessible to users, and aligned with business needs. The criteria of his success are his KPIs: usage, technical performance, data quality", adds Killian.

The last prerequisite is that the business be able to clearly and continuously define its data domains and, once the model has proven its value, be capable of scaling up.
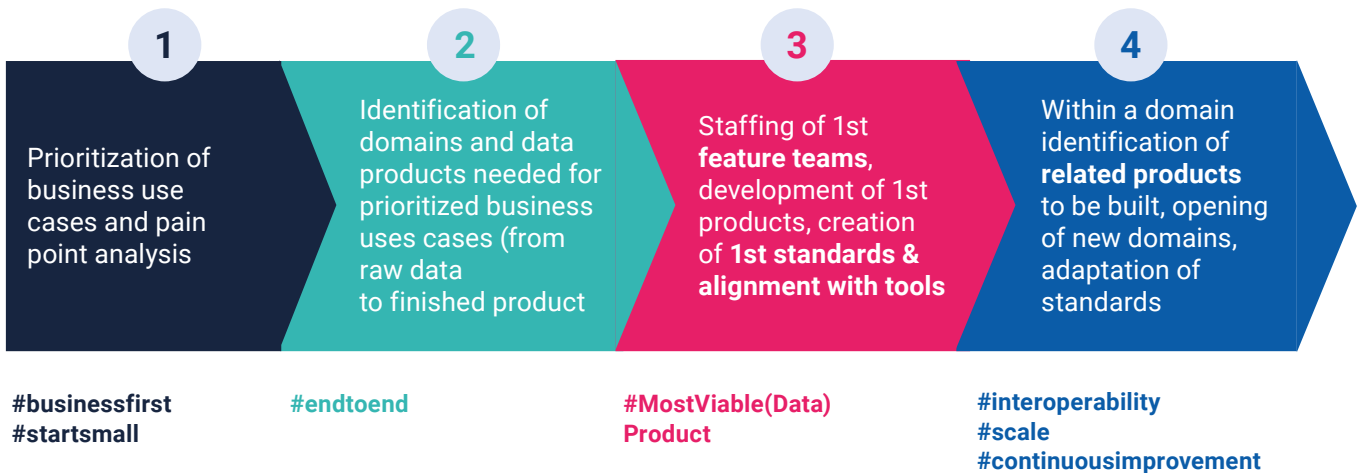
These are the three of the most frequently-asked questions by clients implementing data mesh, along with Artefact's recommendations for successfully defining domains, measuring success, and knowing when it's opportune to scale up.

The tech stack: managing data mesh with Google Cloud

"The first thing data and IT teams need to implement data mesh is the ability to make their data discoverable and accessible by publishing it in a data catalog", begins Amine Mohktari. "To achieve this, Google has a first pillar, Big Query, which enables the creation of shareable datasets. The second pillar, the catalog itself, is

35

The deployment approach used by Artefact clients consists of demonstrating the value of the model on an initial perimeter or domain

1	**2**	**3**	**4**
Prioritization of business use cases and pain point analysis	Identification of domains and data products needed for prioritized business uses cases (from raw data to finished product	Staffing of 1st **feature teams**, development of 1st products, creation of **1st standards & alignment with tools**	Within a domain identification of **related products** to be built, opening of new domains, adaptation of standards

#businessfirst
#startsmall

#endtoend

#MostViable(Data) Product

#interoperability
#scale
#continuousimprovement

made possible by Analytics Hub, which creates links to all the datasets created by various members of the organization or its partners so that subscribers may easily access them."

"It's important to understand that only links to data are made – never copies. Thanks to this system, subscribers can use data as if it belongs to them, even though it remains in its original physical location. This remains true even when you have data sets stored in a different cloud", assures Amine.

User experience is a major principle of the system and is reflected in all aspects of data mesh, not only in facilitating data sharing and data composition, but by keeping data permanently available, no matter how many users are active.

As for data security and governance, Google has it covered with Dataplex,

their intelligent data fabric that helps unify distributed data and automate data management and governance across that data to power analytics at scale. Along with an Identity and Access Management (IAM) framework to assign a unique identity to each data consumer, "Dataplex offers companies a set of technical pillars that allow them to carry out any implementation of governance in the simplest way possible", explains Amine.

"At Google Cloud, our aim is to provide you with a serverless data platform that will allow your data teams to focus on areas such as processes and business use cases, where they have added value no one else can produce." **IMAGE 3**

Google's Dataplex gives users a 360° view of published data products and their quality

Conclusion: three pitfalls to avoid when implementing data mesh

DON'T > Stay stuck in a project vision instead of a product vision

DO > Define priority data products according to different uses;

DON'T > Scale up the new model too rapidly

DO > Test the model with a well-defined operating model;

DON'T > Deploy an overly complex technical ecosystem

DO > Keep the tech stack small to have as many players as possible.

Why is the "data as a product" concept central to data mesh?

Violaine Berland
Data Consulting Director
ARTEFACT

Killian Gaumont
Senior Consulting Manager
ARTEFACT

The data mesh model has been gaining traction in recent years as a way to approach data management in a more modular and decentralized manner. The idea behind a data mesh is to treat data as a product, rather than a by-product (for each use case), and to build data products that are owned and maintained by specific business domains within an organization.

The data product approach: business advantages

One of the main benefits of this approach is that it allows for greater flexibility and agility in how data is used and accessed. Rather than having a centralized team responsible for the ingestion and management of data pipelines, data mesh enables each data domain to be responsible for their own data and to build data products that are tailored to business needs. Decentralizing data ownership reduces bottlenecks: new data sources can be integrated more quickly, and changes to data can be made more easily in response to changes in the business.

Another benefit of this "product-thinking" philosophy is that it encourages collaboration between domains and thus business units. By treating data as a product, it becomes easier for different teams to share and use the data of others, in ways that are meaningful to them. This can also lead to new insights and opportunities

that would not have been possible with a more siloed approach to data management. For example, new data value appears when a product's sales can be compared with its raw materials composition and pricing evolution.

Implementing data as a product: organizational concerns

However, this new approach does come with a few challenges. One of the most important is how to manage the dependencies between different data products. As each business domain is responsible for their own data, changes made by one domain can have an impact on others. This means clear policies and coordination between domain teams must be established to ensure that data is being used and managed consistently. This is why a central office must coordinate the implementation of "federated governance principles". Another challenge is ensuring that data is of sufficient quality and trustworthiness

to be used as a product. This requires a strong focus on data governance and quality control, as well as regular monitoring and testing of data products to ensure that they are meeting the needs of the business.

Data mesh: a strategic approach for data management

Despite these challenges, the data mesh approach has the potential to bring significant value to organizations. By treating data as a product, and by building data products that are tailored to the specific needs of different business objectives, organizations can unlock new insights and opportunities that would not have been possible with a more siloed approach to data management. With the right governance and quality control in place, the data mesh approach can help organizations to navigate the ever-increasing volume of data and to turn that data into a strategic asset for the business.

AI Industry Solutions

Demand Forecasting

Demand forecasting: Using machine learning to predict retail sales

Jérôme Petit
Managing Partner
Retail & eCommerce
ARTEFACT

Pascal Coggia
CEO
ARTEFACT UK

All industries aim to manufacture just the right number of products at the right time, but for retailers this issue is particularly critical as they also need to manage perishable inventory efficiently.
Too many items and too few items are both scenarios that are bad for business. (Estimates suggest that poor inventory management costs US retailers close to two billion dollars per year.)

Looking beyond past sales to accurately predict future sales

Massive incremental profit can be unlocked by retailers managing orders and inventory effectively. But as this requires the processing of data for a huge number of stock keeping units (SKUs), which often include perishable goods and items that are ordered daily, it is also a significant challenge.

Retailers used to rely solely on the data from previous years to predict future sales (and therefore manage their inventory), but this method is only useful up to a point. However, machine learning has now evolved to the stage that it can provide accurate predictive models using different signals based on how they influence purchases.

Predicting sales is complex because, in any given period, purchases are affected by many factors: weather, shopping trends, regulation, new products, buying behaviours, a pandemic… And predictions based on previously recorded data don't factor in specific events, making monthly sales appear evenly distributed when this is unlikely to be the case.

For example, an item that is often out of stock might cause a slowdown in the sales of that particular product or category, but it won't show in the monthly reports. Even worse, poor figures are often regarded as a mark of buyers' disinterest, when the opposite is true; consumers' over-purchase of an item has caused it to sell out.

Or a product missing from the store might actually be in stock – just not yet out on the shelves. Big box retailers often struggle to restock in real time, so an instantly popular item might disappear from the shelves very quickly, and thus not perform as well as expected, despite it being available in inventory. This calls for technology that can help retailers seamlessly align supply and demand.

Using machine learning and multiple signals to assess inventory levels

Machine learning provides a solution to these challenges. Predictive models can forecast sales months in advance by using a number of the signals that affect them (seasonality, consumption trends, price levels, etc). To be as accurate as possible, it's important that the models use more indicators than the standard day, product and store that it is usual to factor in.

To illustrate this, a retailer might analyse the seasonality to predict sales for the forthcoming period. However, the data will be skewed because using dates is not 100% accurate; a certain date can be a weekday one year, but the weekend the following year, causing sales to vary greatly. Other factors, such as whether that date falls on a holiday (Christmas, Easter, etc) or a major sporting event, also influence consumer buying patterns.

It is a similar story with price level signals. Promotions at store level can markedly affect the sales of a product from a given category or even make the store as a whole more attractive.

Both these examples illustrate why it is necessary to take many different signals and indicators into account to accurately forecast sales: a task that was a pain in the neck before machine learning and advanced artificial intelligence models made it achievable.

Adopting machine learning for inventory management

The technology is there, but for retailers to use it effectively and make accurate predictions, they need to collect and analyse huge amounts of data. Much of this is in different data sources and it can be complex to try to process multiple Excel and PDF files that contain previous reports and media plans. Big data tools are needed to process this information into the clean and readable format required to create predictive models that can prevent inventory issues.

Past sales data for a given store may be 'inaccurate' due to one-off events (promotions, adverse weather, traffic congestion, etc). To remove this bias, predictive models combine past sales numbers with those of similar stores.

The other big challenge is preventing items being unavailable on the shelves while they are in stock (caused by it being almost impossible for employees to monitor shelves in real time and restock them immediately).

Technology solutions using surveillance cameras and weight sensors do exist but are a huge investment. However, readily available information such as real-time sales at SKU level may be leveraged to detect 'empty shelf' situations. Models can analyse the usual flow of sales of an item, so the normal time between two sales of a product in a given store is known. Human intervention can be used to review and resolve statistical anomalies.

Predictive analysis is but one of the many ways that traditional retailers can benefit from machine learning. They have a lot to gain from relying on advanced technology for better inventory management to increase store revenue. Processing vast amounts of data can also help them to optimise the assortment, offer more attractive and profitable promotions and set prices more efficiently.

Well-devised tools can undertake complex and time-consuming tasks and quickly deliver accurate reports. This is the real value creation lever of artificial intelligence in retail: freeing managers from tedious multi-sources comparisons analyses and allowing them to focus on the continued improvement of the customer experience.

L'ORÉAL Trend detection

Innovating tomorrow's products today thanks to AI trend detection by Artefact

Charles Besson
Global Social Insights & AI Director - L'ORÉAL

Fabrice Henry
Managing Partner - ARTEFACT

Charles Besson, Global Social Insights & AI Director at L'Oréal, and Fabrice Henry, Managing Partner at Artefact, discuss how L'Oréal Trend Detection, deployed with Artefact's AI trend detection solution, is predicting what cosmetics products consumers are going to want tomorrow.

CHALLENGES

Predicting new trends before the competition.

L'Oréal is the world's leading beauty company, present in 150 countries, offering a rich portfolio of iconic brands for every type of consumer. The company's socially responsible programme, Sharing Beauty With AI, is dedicated to shaping the future of beauty through major, sustainable product innovations.

"There's a quest at L'Oréal to constantly reinvent the business, the brands… it's part of our DNA, it's an obsession,"

says Charles Besson,
Global Social Insights & AI Director at L'Oréal.

L'Oréal has great tools for product innovation: a Prospective Consumer Intelligence department, a Digital department,

an IT department. Their latest programme, the L'Oréal Beauty Tech accelerator, enables the company to select the most promising innovative products for incubation and acceleration. But how to make L'Oréal the world's number one beauty tech company? Charles wondered:

"Given the abundance of public data we can collect, and using this extraordinary thing that is artificial intelligence, could we invent an algorithmic crystal ball capable of predicting the future?"

After consulting with several other companies, he chose Artefact to build his dream machine.

41

ARTEFACT

L'ORÉAL

SOLUTION

A co-creation that can forecast emerging consumer trends.

L'Oréal's ambitious project needed to go a step further with AI technology, in comparison with traditional market research models. That's where Artefact leveraged its advanced expertise in digital marketing and data science to help L'Oréal detect and predict new trends emerging in the digital space. Because discovering what consumers want – almost before they know they want it – is the Holy Grail every marketer seeks.

Developing an innovative and reliable trend prediction solution was both exciting and challenging for the project team. As soon as they started brainstorming, they realised that tracking influencers wasn't the answer.

"Sure, when Kim Kardashian wears a new lipstick, everyone starts buying the same colour, but by then it's already too late, the million-dollar question is what happens before that?",

explains Fabrice Henry, Managing Partner at Artefact.

So we went deeper, and asked upstream questions to find out where trends originate and how they propagate. Once a trend is born, how does it spread? Does it spread differently according to geography or community? What are the big sources – YouTube, blogs, Instagram, Facebook, etc. – from which data can be extracted in order to train algorithms?

"We found different approaches for each of these subjects and proposed a final one to Charles Besson. And that's where our collaboration began, where we started this project," adds Fabrice.

The project co-created by L'Oréal and Artefact was based on three key success factors: co-development of an employee-centric solution, validation of the solution via an MVP (Minimum Viable Product) prior to scaling, and strong collaboration based on trust – a vital element when sharing sensitive information with your partners.

RESULT

This project is a predictive intelligence machine with three main components:

- **DETECT**: Using Natural Language Processing (NLP) algorithms, this feature can digest a database composed of millions of documents and extract weak signals – keywords that are relevant but rare (e.g. emerging terms) in the beauty domain

- **PREDICT:** Once new, atypical, or relevant beauty terms have been detected, we have to see if they have staying power. To find out, we train machine learning algorithms based on predictive variables that have reliably demonstrated whether a given trend is going to grow or not, using factors such as number of mentions, commitment score, co-occurrence of author citations, etc."

- **ILLUSTRATE**: Building a number of visualisations to demonstrate the power of the trend, along with a variety of contextual elements (brands or authors that were talking about it, articles and visuals that mentioned it…) and let all of this appear in the tool's interface.

"I'm really happy with it, we've launched a beta version, and if all goes well; the next steps will be the launch, adoption, and training. We've already had lots of positive feedback!"

concludes Charles Besson, Global Social Insights & AI Director at L'Oréal.

Scoring customer propensity using machine learning models on Google Analytics data

Antoine Aubay
Data Science Manager
ARTEFACT

A deep-dive on how we built state of the art custom machine learning models to estimate customer propensity to buy a product using Google Analytics data.

- Propensity modeling can be used to increase the impact of your communication with customers and optimize your advertising budget spendings.

- Google Analytics data is a well structured data source that can easily be transformed into a machine learning ready dataset.

- Backtest on historical data and technical metrics can give you a first sense of your model's performance while live test and business metrics will allow you to confirm your model's impact.

- Our custom machine learning model outperformed existing baselines: during live tests in terms of ROAS (Return on advertising spend): +221% vs rule based model and +73% vs off-the-shelf machine learning (Google Analytics session quality score).

This article assumes basic fundamentals in machine learning and marketing.

What is propensity modeling ?

Propensity modeling is estimating how likely a customer will perform a given action. There are several actions that can be useful to estimate:

- Purchasing a product
- Churn
- Unsubscription
- etc …

In this article we we will focus on estimating the propensity to purchase an item on an e-commerce website.

But why estimate propensity to purchase ? Because it allows to adapt how we want to interact with a customer. For exemple, suppose we have a very simple propensity model that classify the customers in "Cold", "Warm" and "Hot" for a given product ("Hot" being customers with highest chance of buying and "Cold" the least)
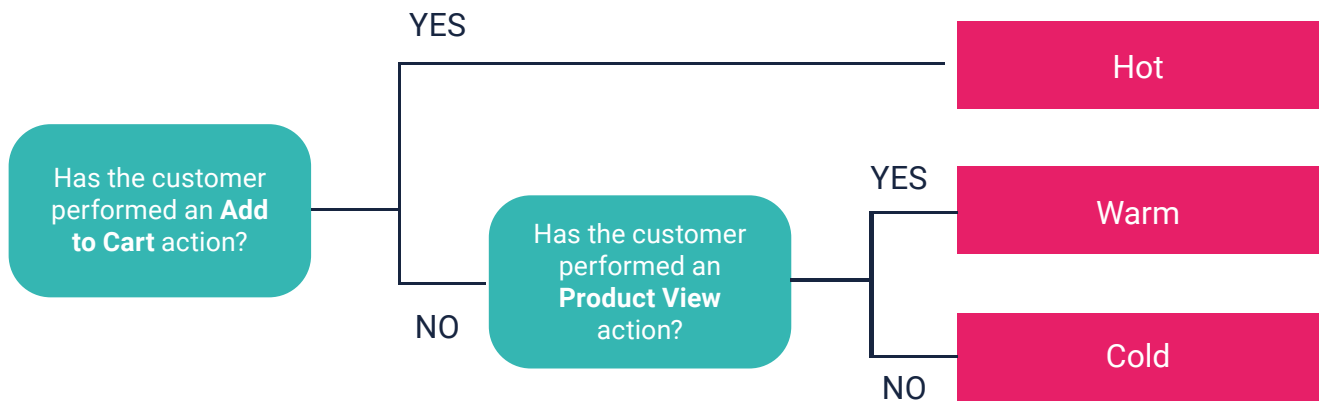
Well, based on this classification you can have a specific targeted response for each class. You might want to have a different marketing approach with a customer that is very close to buying than with one who might not even have heard of your product. Also if you have a limited media budget , you can focus it on customers that have a high likelihood to buy and not spend too much on the ones that are long shots.

Simple Rule Based Propensity Model

This simple type of rule based classification can give good results and is usually better than not having any but it has several limitations:

• It is likely not exploiting all the data you have at your disposal whether it be more precise information on the customer journey or your website or other data sources you may have at your disposal like CRM data.

• While it seems obvious that customers classified as "Hot" are more likely to purchase than "Warm" which are more likely to purchase than "Cold", this approach does not give us any specific figures on how likely they are to purchase. Do "warm" customers have 3% chance to purchase ? 5%? 10% ?

• Using simple rules, the number of classes you can obtain is limited, which limits how customized your targeted response can be.

To cope with those limitations we can use a more data driven approach: use machine learning on our data to predict a probability of purchase for each customer.

Understanding Google Analytics data

Google Analytics is an analytics web service that tracks usage data and traffic on website and applications.

Google Analytics data can be easily exported to Big Query (Google Cloud Platform fully managed data warehouse service) where it can be accessed via an SQL like syntax:

Note that the Big Query export table with Google Analytics data is a nested table at session level:

• Sessions are a list of actions a specific customer does within a given timeframe. They start when a customer visit a page and end after 30 minutes of activity.

• Each customer can have several sessions.

• Each session can be made of severals hits (i.e. events) and each hit can have a several attributes or custom metrics (this is why the table is nested, for instance if you want to look at a the data at hit level you will need to flatten the table).

For example in this query we are only looking at session level features:

```
SELECT
VisitId,
fullVisitorId,
totals.hits,
totals.pageviews,
totals.timeOnSite,
device.browser,
geoNetwork.country,
device.operatingSystem,
channelGrouping
```

```
FROM'bigquery-public-data.
google_analytics_sample.ga_
sessions_20170801'

WHERE totals.hits > 1
```

And in this query we have used an Unnest function to query the same information at hit level:

```
SELECT
VisitId,
hits.hitNumber,
hits.page.hostname,
hits.page.pagePath,
Hits.evenInfo.eventAction

FROM'bigquery-public-data.
google_analytics_sample.ga_
sessions_20170801'
UNNEST(hits) as hits

WHERE totals.hits > 1
```

Note that our project was developed on GA360 so if you are using the latest version, GA4, there will be some slight differences in data model, especially the table will be at event level. There are public sample tables of GA360 and GA4 data available on Big Query.

Now that we have access our raw data source we need to perform feature engineering before we can feed our table to a machine learning algorithm

General Features

- Total number of sessions
- Total number of hits
- Bounce Rate
- Total Add to cart
- Total page views
- Recency since last session

Favorite Features

- Favorite browser
- Favorite operatingSystem
- Favorite channelGrouping
- Favorite region
- Favorite deviceCategory

Product Features

- Number of page views of product
- Number of add to cart of product
- Total hits with interactions on the product
- Total sessions with at least one interaction with product
- Recency since last session with at least one interaction with product

Similar Product Features

- Number of page views of similar products
- Number of add to cart of similar products
- Total hits with interactions on similar products
- Total sessions with at least one interaction with similar products
- Recency since last session with at least one interaction with similar products

Crafting the right features

The aim of the feature engineering step is to transform the raw Google Analytics data (extracted from Big Query) into a table ready to be used for Machine Learning.

GA data is very well structured and will require minimal data cleaning steps. However there are still a lot of information present in the table, many of which are not useful for machine learning or cannot be used as is so selecting and crafting the right features is important. For this we built features that seemed to be the most correlated with buying a product.

We crafted 4 types of features:

Features used in the model

Note that we are computing all those features at a customer level which means that we are aggregating information from multiple sessions for each customer (using fullVisitorId field as a key)

GENERAL FEATURES

Global features are numerical features that give general information about the session.

Note that bounce rate is defined as % of times the customer only visited only one webpage during a session.

It was also important to include information on the recency of events: for instance a customer that just visited your website is probably more keen to purchase than one that visited it 3 months ago. For more information on this topic you can check the theory on RFM (recency, frequency monetary value).

So we added a feature Recency since last session = 1 / Number of days since last session which allows the value to be normalized between 0 and 1

FAVORITE FEATURES

We also wanted to include some information on the key categorical data available such as browser or device. Since that information is at session level, there can be several different values for a single customer so we only take the one that occurs the most per customer (i.e. the favorite). Also, to avoid having categorical features with too high cardinality, we only keep the 5 most common values for each feature and replace all the other values with an "Other" value

PRODUCT FEATURES

While the first two types of features are definitely useful in helping us answer the question "Is a customer going to buy on my website?", they are not specific enough if we need to know "Is the customer going to buy a specific product?". To help answer this question we built product specific features that only include the product for which we are trying to predict the purchase:

For Recency since last session with at

least one interaction with this product, we use the same formula than for the Session Recency in the General Features. However we can have cases where there is 0 session with at least one interaction with the product, in which case we fill with 0. This makes sense from a business perspective since is our highest possible value is 1 (when the customer had a session since yesterday).

SIMILAR PRODUCT FEATURES

In addition to looking at the customer's interaction with the product for which we are trying to predict the probability to purchase, knowing that the customer interacted with other products with similar function and price range can definitely be useful (ie substitute product). For this reason we added a set of Similar Product features that are identical to the Product features except that we also include similar products in the variable scope. The similar products for a given product were defined using business inputs.

We now have our feature engineered dataset on which we can train our machine learning model.

Training the model

Since we want to know whether a customer is going to purchase a specific product or not, this is a binary classification problem.

For our first iteration, we did the following to create our machine learning dataset (which was 1 row per customer):

• Compute the features using the sessions in a 3 months time window for each customer.

• Compute the target using the sessions in a 3 weeks time window subsequent to the feature time window. If there is at least one purchase of the product in the time window, Target it equal to 1 (defined as Class 1), else Target is equal to 0 (defined as Class 0)

• Split the data between a Train set and a Test using 80 / 20 random split.

However some first data exploration quickly showed that there was a strong class imbalance issue: Class 1 / Class 0 ratio was over 1:1000 and we did not have enough Class 1 customers. This can be very problematic for machine learning models.

To cope with these issues we made several modifications in our approach:

• We switched the target variable from making a purchase to making an add to cart. Hence, our model looses a bit in terms of business signification but increasing the volume of Class 1 more than compensates.

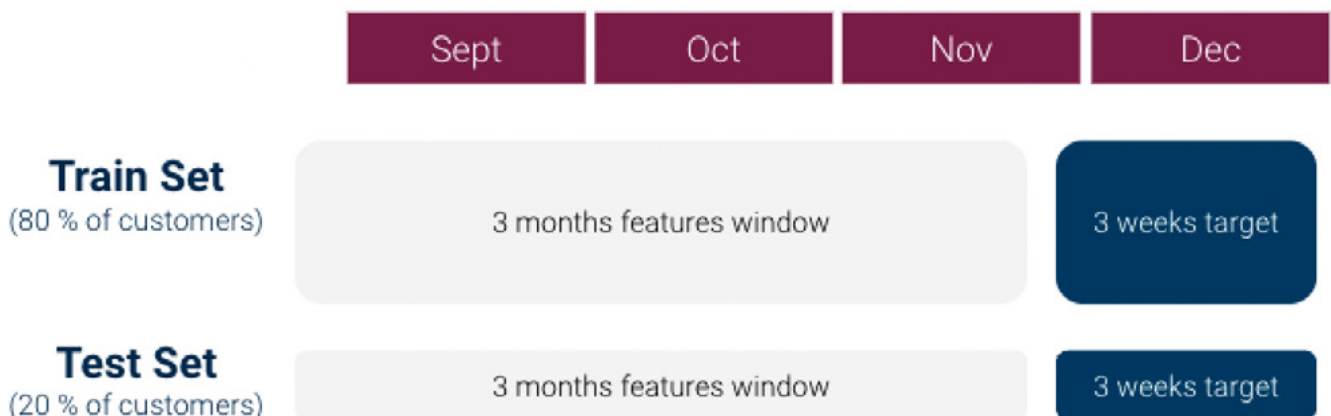• We trained the model on several shifting windows, each of 3 months + 3 weeks, instead of a single one. In addition to increasing our volumes of data, this improves the generalization capacity of the model by training on various time of the year where the customers can have different purchase behaviors. Note that due to this, the same customer will be present several times in the dataset (at different periods). To avoid data leakage we make sure that he is always either in the training or the test dataset.
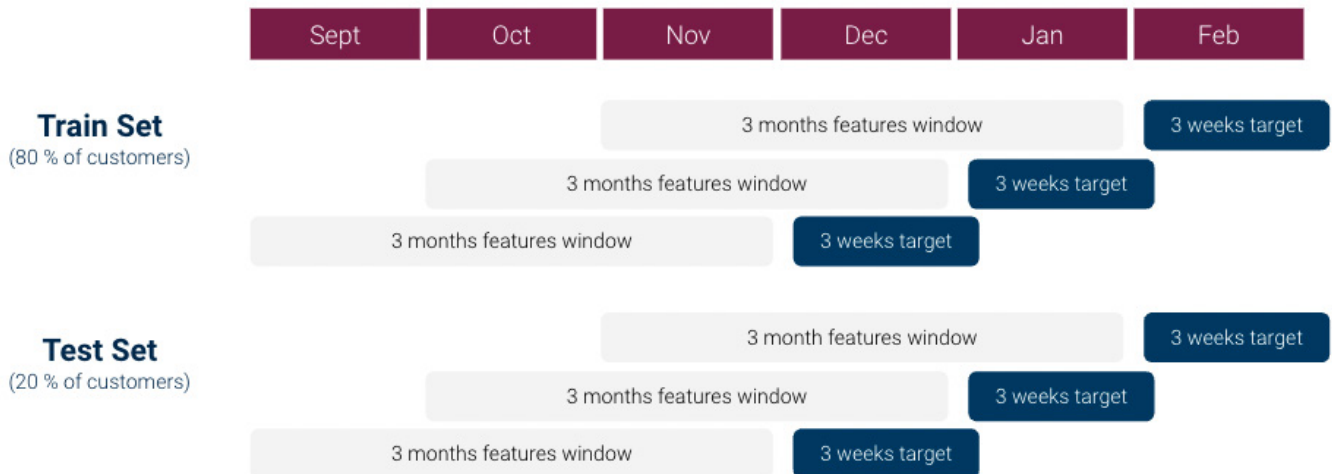
• We undersampled our Class 0 so that the Class 1 / Class 0 ratio is 1. Undersampling is a good solution to deal with the class imbalance issue, compared to other options such as oversampling or SMOTE, because we were already able to increase the volume of Class 1 considerably with the first two changes. Only the training set is rebalanced since we want the test set to have the same class ratios than the future data we will test it on. Note that we tested with higher ratios such as 5 or 10 but 1 was optimal in model evaluation.

Using this dataset we tested with several classification models: Linear Model, Random Forest and XGboost, finetuning hyperparameters using grid search, and ended up selecting an XGboost model.

Evaluating our model

When evaluating a propensity model there are two main types of evaluations

	Sept	Oct	Nov	Dec
Train Set (80 % of customers)	3 months features window			3 weeks target
Test Set (20 % of customers)	3 months features window			3 weeks target

that can be performed:

• Backtest Evaluation
• Livetest Evaluation

BACKTEST EVALUATION

First we performed backtest evaluation: we applied our model to past historical data and checked that our model is correctly identifying customers that are going to perform an add to cart. Since we are using a binary classifier, the model produces a probability score between 0 and 1 of being Class 1 (Add to cart).

A first step in evaluating a binary classification model is create a confusion matrix and compute the precision / recall (or their combined form in the f1 score). However there are two issues with these simple metrics:

Some can be hard to interpret because the dataset is imbalanced (for instance the precision metric will generally be very low because we have so few Class 1)

They require to decide on a probability threshold to discriminate between Class 0 and 1

Confusion Matrix Example for our Class Imbalanced problem

So we decided to use two metrics that were more interpretable:

• PR AUC: Area under the curve of precision by recall graph (see this explanation for more details). Essentially this metric allows us to get a global evaluation on every possible threshold. This metric is well suited for unbalanced dataset where the priority is to maximize precision and recall on the minority class : Class 1(contrary to its cousin the ROC AUC)

• Uplift: we sort customers by their probability score and we divide our results into 20 ventiles. Uplift is defined as the Class 1 Rate in the top 5% / the Class 1 Rate across all the dataset. So for instance if we have 21 % Add to Cart in the top Top 5 % of the dataset vs 3 % Add to cart Rate in whole dataset we have an uplift of 7 which means our model is 7 times more effective than a random model.

Results on those metrics were rather positive, especially, Uplift was around 13.5.

Backtest evaluation is a risk free method for a first assessment of a propensity model but it has several limitations:

Since it is only done on the past, the model output is not actually being

used to impact the media budget strategy.

With our metrics, we only assessed if the model was able to correctly identify customers who would make an add to cart but we did not assess how the identification of those customers would generate a sales uplift.

LIVETEST EVALUATION

So to get a better idea of our model's business value we need to perform live test evaluation. Here we activate our model and use it to prioritize advertising budget spendings :

Results we obtained on the livetest were very solid:

• Compared to a simple rule based approach for evaluation propensity, our model's ROAS was +221 %

• Furthermore we also compared our performance to a strong contender competition in the form of Google's Session Quality Score: a score provided by Google in the Google Analytics dataset, and in that case our model was still at +73 % ROAS. This shows how a custom ML approach can bring considerable business value.

	Predicted as Class 0	Predicted as Class 1	
Class 0 (No add to cart)	True Negative : 1 500 000	False Positive : 100 000	**Recall** = TP / (TP + FN) = 83 %
Class 1 (Add to cart)	False Negative : 150	True Positive : 750	**Precision** = TP / (TP + FP) = 0.7 %
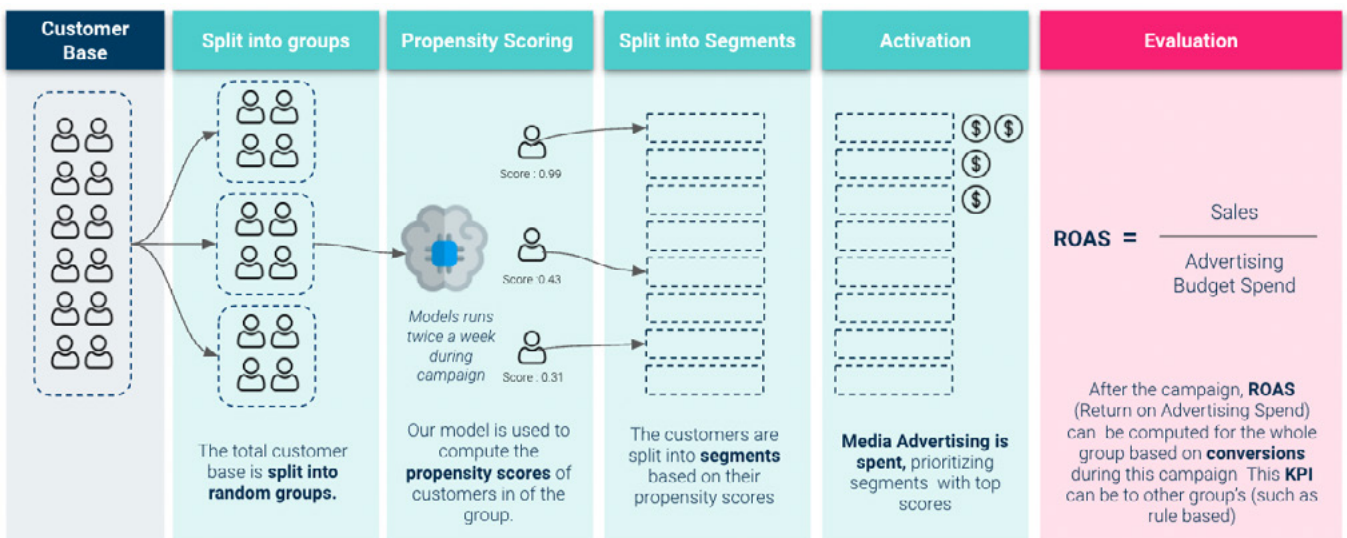
Conclusion

In addition to reaching solid performance, a strong side benefit of our approach is that our feature engineering is very generic. Almost none of the feature engineering steps need to be adapted to apply our model to a different country scope or product scope. In fact following our first success in the livetest, we were able to roll out our model to multiple countries and products in a very efficient manner.

AI Industry Solutions

AI for Call Centre

ARTEFACT

Powering your call centre with artificial intelligence

Matthieu Myszak
VP Data Consulting
ARTEFACT

In today's competitive business environment, customer experience is becoming a key differentiator for organizations well aware that it is more cost-effective to keep an existing customer than acquiring a new one and that a disgruntled customer well-handled can be turned into an advocate for your brand.

The recent advancements in technology and artificial intelligence can be applied to your own customer interactions and be a tremendous support to help your organization profit from productivity gains, improve customer retention and create additional revenue.

Having your customers talk to a robot can seem out of a scifi novel. But actually, it is already a reality for hundreds of millions of customers that have regular interactions with interfaces powered by artificial intelligence.

Ultimately, the goal of an optimal integration is to have customers talk like they usually do when interacting with a conversational agent that can analyse their sentiment, provide useful information, and answer recurrent standard requests as well as complex issues. Also, the virtual robot can pass on the caller to a human agent when needed. Beyond this conversational agent, Machine Learning technologies record the interaction for further improvement in setting new and sophisticated protocols of problem solving.

1 — The main advantages of using artificial intelligence for call centres

What differentiates customer service boils down to the quality of the relationships and the overall user experience. In that regard, AI can be a useful tool helping organizations achieve unmet customer expectations.

An artificial intelligence-powered customer centre is a crucial asset for three reasons:

- Reduce operating costs (via automatisation, reduction of average handle time)

- Improve the quality of the service and the customer satisfaction (via an increased reactivity and availability)

- Offer opportunities for cross-selling and upselling customers

The revolution of conversational technologies has already begun. Today, internet users are becoming more and more familiar with voice commands and chatbot interactions thanks to natural language understanding capabilities (NLP) wired into mainstream digital products, such as Google Voice Search, Google Assistant or Google Home.

According to a recent Gartner study, by 2023, 70% of consumers will prefer to interact with a vocal interface than a real person and 40% of all customer interaction will be fully automated.

2 — How does it work

Google Cloud has been developing virtual assistants capabilities for years and has created a product that can be used for business purposes.

Artificial intelligence is not always meant to replace humans as it can be utilized to augment real agents. Thanks to the robot, call centre operators can concentrate on complex and higher value situations and be freed from small, repetitive and low-value tasks. The chatbot is also critical in providing assistance to the operator. In certain situations, maintaining a human contact is key.

With the support of artificial intelligence, the customer service agent becomes an "augmented agent", meaning that the virtual assistant listens to calls in real-time and provides contextual assistance letting the human stay focused on the conversation and expressing empathy towards the customer.

3 — Concrete business cases showing tangible benefits in improving customer satisfaction

Businesses of all sorts could benefit on both sides from a positive impact by powering their customer service with artificial intelligence:

FOR CALL SERVICE OPERATORS

Improve efficiency and focus, reduce churn, provide opportunities to upsell and cross-sell

FOR CUSTOMERS

Improve user experience with a customer service available 24/7 with no waiting time, accelerate retention and customer loyalty

"If a customer asks for the status of his order, the virtual assistant has to provide the right request and give the correct information."

4 — Seamless integration into your legacy system

For optimal performance, the Google Contact Centre AI must be integrated into the call centre workflow, work with the existing databases and documentation (via APIs) and the front desk interfaces.

Organizations need to bring a multidisciplinary team to achieve this project according to their needs and their own IT architecture.

Before being set up, a chatbot needs to be fed with customer interaction data. The bot needs to be trained by listening and analyzing past customer interactions. That will enable the virtual assistant to be able to provide value immediately with high levels of customer satisfaction. Existing data can be emails, chat messages or voice calls. The data will help train the model according to the customer journey and the expected optimisations.

"If we have to manage use cases of a client from the banking industry, we will not train the AI tool the same way that if it was for an ecommerce brand, for example."

It can take from one up to three months to integrate the artificial intelligence solution into an existing customer service depending on the number and the complexity of use cases, and number of integration points to access.

"Deploying a virtual agent for an insurance company that can automatically file a damage claim is more complex for example than if you are asking for the status of your order on a ecommerce website."

In the event that a business doesn't have any data to analyze or precise use cases to aim for, it is possible to implement a working solution by asking each caller a prompt such as "Could you please tell us the reason for your call ?" and then letting them access the traditional customer experience journey. By analyzing the initial answer and the human agents' interactions, the artificial intelligence will get trained to quality future interactions.

5 — Why rely on a partner

Before taking the task of implementing a virtual assistant solution into your architecture, it could be useful to bring in an experienced partner that could assist you in the different steps of the project and help you maximise value.

Artefact has been helping clients, in various industries, turbocharge their call centres with artificial intelligence. We provide assistance in different ways:

• Identification and prioritization of use cases

• Setup and training of artificial intelligence solutions

• Development of integrations to collect relevant data

Our company has extensive experience working with both partners and service providers from the digital, data and artificial intelligence industries. Artefact's method is centreed on featured teams, composed of members with complementary skills, from business

analysts to data scientists and software engineers, that can help projects come to life.

"We don't usually think of adding a UX Designer to feature teams but this role is important, as it helps give a personality to the virtual agent that reveals your unique brand image"

Our company has extensive experience working with both partners and service providers from the digital, data and artificial intelligence industries. Artefact's method is centered on featured teams, composed of members with complementary skills, from business analysts to data scientists and software engineers, that can help projects come to life.

Setting up a bot that is both well-embedded into your data architecture and provides usefulness to consumers is a goal that can take some effort but can be reached when organizations make it a customer experience priority.

Conclusion

An artificial intelligence, such as Google Contact Centre AI, integrated into the data architecture of your customer service will supercharge your customer experience.

The machine learning capabilities and recordings of interactions provide a constant feedback loop that helps the performance of the virtual assistant and the use cases that can be addressed.

The right partner for an artificial intelligence project can help organizations smooth out the definition and implementation phases and achieve immediate performance and business gains, while providing useful industry benchmarks and key learnings from previous experiences.
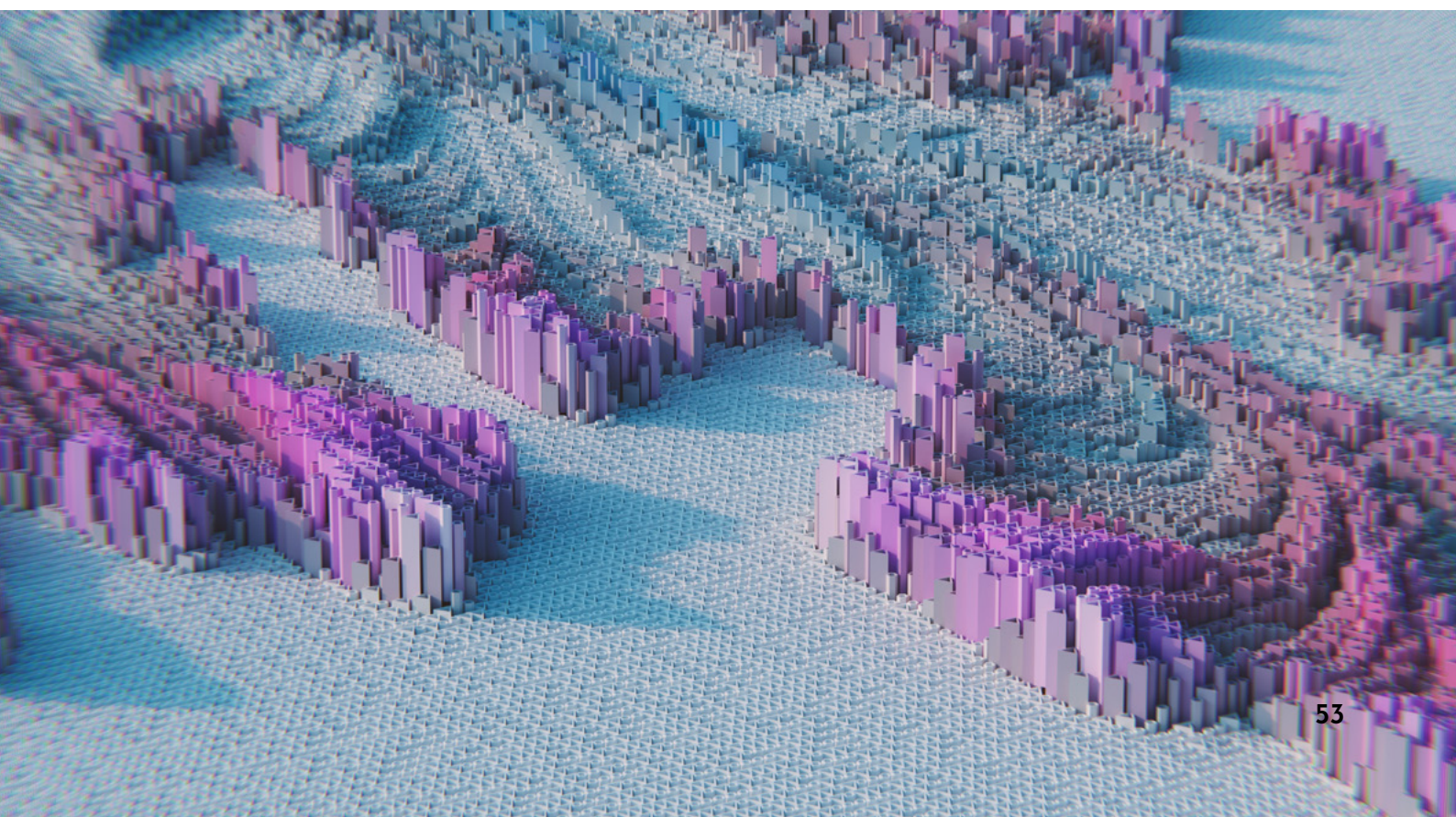
MAIF

CASE STUDY

MAIF
Using topic modelling to reduce contact centre bottlenecks

CHALLENGES

MAIF is one of France's largest home and automotive insurance companies, with more than **3 million members.**

One of the challenges facing its customer services team was managing the volume of calls coming into its call centre — on average, **some 8 million a year.**

With no way of vetting calls before they reached an operator, the team was wasting precious time responding to questions customers easily find the answers to on the MAIF website.

To improve efficiencies, we needed to filter out unnecessary calls and free up more time for MAIF's customer service teams to process more complicated requests.

SOLUTION

To understand why customers were calling MAIF's call centre, we developed **Natural Language Processing** (NLP) algorithms to analyse transcripts of more than 4 million calls.

We then used **topic modelling to categorise every call** into one of 35 different request types.

We liaised with MAIF's business teams to identify which questions could be solved online and which needed a human response or presented a sales opportunity.

Where calls did not represent an opportunity, we advised how to answer these questions **online.**

RESULTS

Our analysis showed that 32% of inbound calls were 'low added value requests' — questions that could easily be answered online.

As a result, we built a roadmap advising MAIF how to solve these questions online and direct people to this content to avoid calling.

Digitising these queries has let MAIF's customer services team prioritise cases that require a human touch, improving efficiencies and its round-the-clock service.
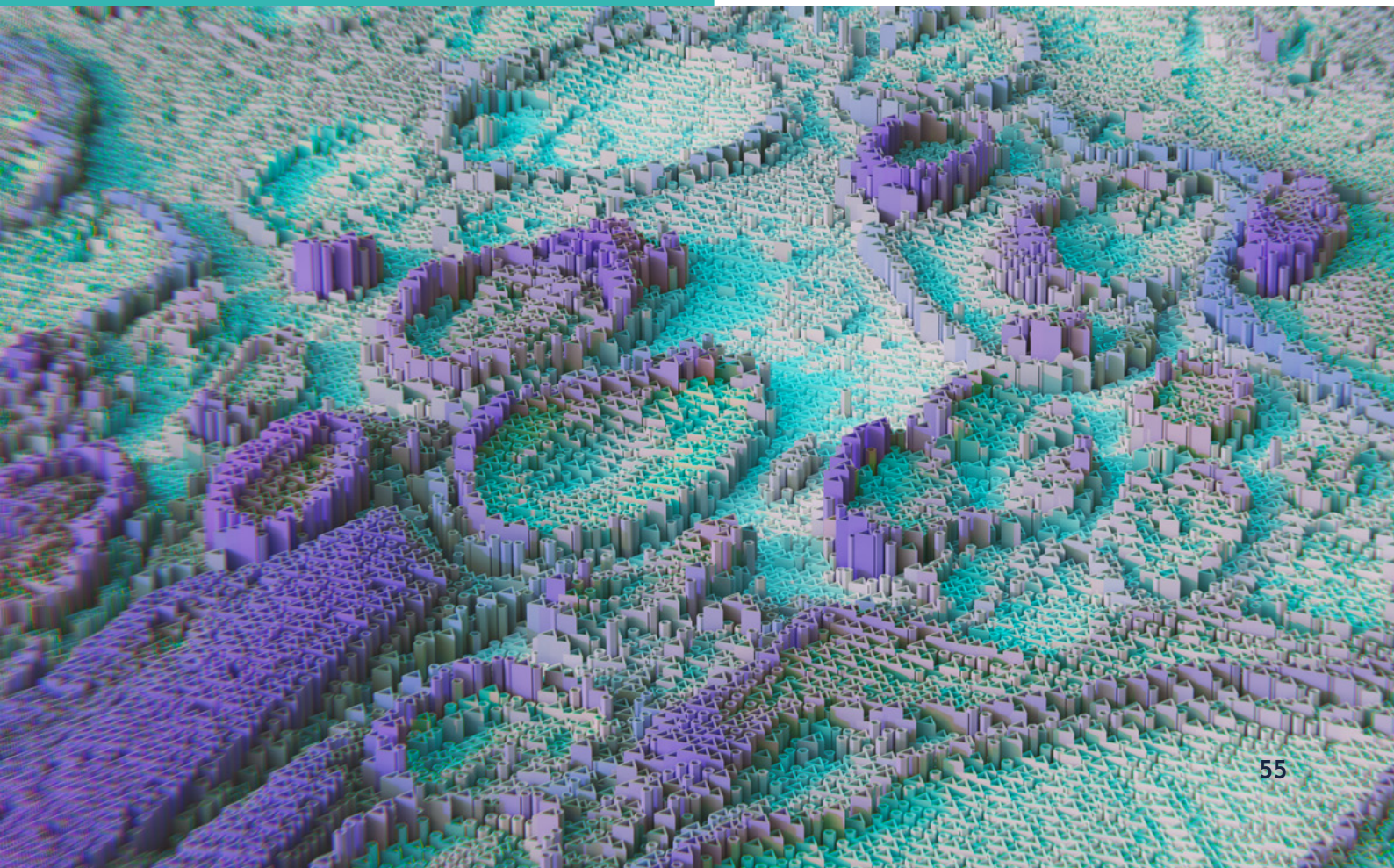
Using NLP to extract quick and valuable insights from your customers' reviews

Everyone talks about BERT, GPT-3, XLNet… but did you know that with some simple NLP 101 preprocessing you can already extract valuable insights from your data?

Louise Morin
Senior Data Scientist
ARTEFACT

Understanding customers' feedback and knowing what your strengths and weaknesses are is key to any business. Nowadays, companies have access to a lot of information that could give them those insights: website reviews, chat interactions, conversations transcripts, social media comments…
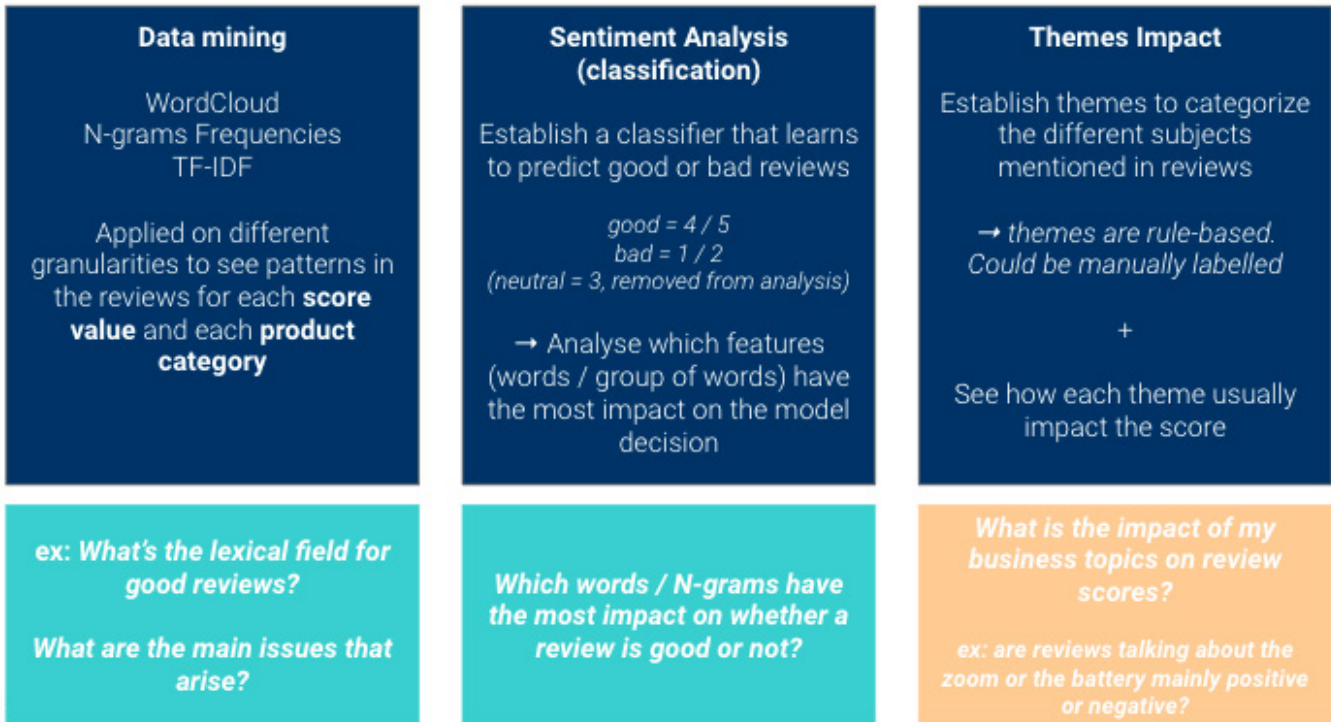
This article explains how you can quickly extract insights from textual data, leveraging consumers' reviews as an example. We will present 3 different approaches:

- Unsupervised data exploration

- Sentiment analysis with features importance

- Analyzing correlation between ratings and predefined business themes

(topic modeling could be a fourth option to go further)

Please note the data behind this article was artificially generated to ensure confidentiality of our initial project.

Data mining

WordCloud
N-grams Frequencies
TF-IDF

Applied on different granularities to see patterns in the reviews for each **score value** and each **product category**

ex: What's the lexical field for good reviews?

What are the main issues that arise?

Sentiment Analysis (classification)

Establish a classifier that learns to predict good or bad reviews

good = 4 / 5
bad = 1 / 2
(neutral = 3, removed from analysis)

→ Analyse which features (words / group of words) have the most impact on the model decision

Which words / N-grams have the most impact on whether a review is good or not?

Themes Impact

Establish themes to categorize the different subjects mentioned in reviews

→ *themes are rule-based. Could be manually labelled*

+

See how each theme usually impact the score

What is the impact of my business topics on review scores?

ex: are reviews talking about the zoom or the battery mainly positive or negative?

Customer Reviews Analysis

We are trying to find insights from our products reviews in order to understand what are their main issues / main strengths. Products are camera devices and accessories, rated from 1 (bad) to 5 (excellent).

We will be using three different approaches here, to gather insights from our data.

The point is to have complementary views:

- Data mining or sentiment analysis is more exploratory: it will find out what matters the most, what could be the main reasons driving a review to be positive or negative.

- Themes impact is used to associate scores distribution to already defined business concepts (zoom, battery, …).

Number of reviews per product category
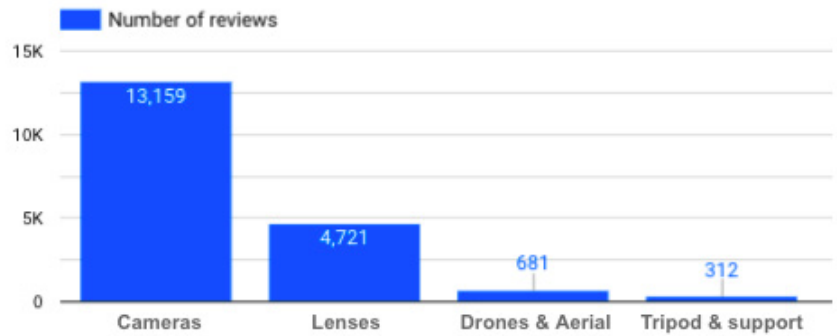
Get a global look at the data you have collected

Whenever you're starting a new data project, the first step is always to get the global picture on the data you have (is it imbalanced? is there enough data? are there lot of missing values?).

HOW MANY REVIEWS DO I HAVE FOR EACH PRODUCT CATEGORY?

The fact that there are not as many Tripod reviews should be kept in mind if we analyze reviews for this specific category of product. The more data we have, the better, in order to have unbiased and relevant conclusions.
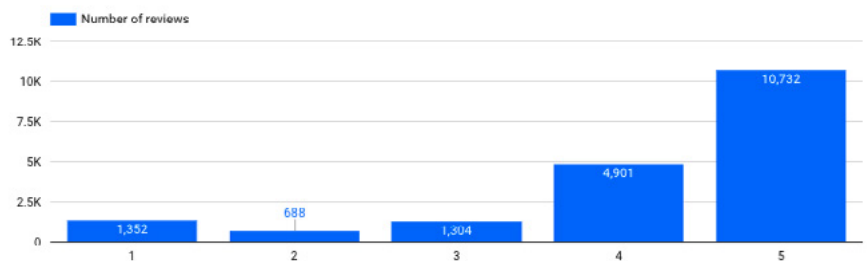

Number of reviews per score

HOW MANY REVIEWS DO I HAVE FOR EACH RATING?

This is important. We see that our dataset is quite imbalanced, we have a lot more positive reviews than negative reviews. This kind of information needs to be taken into account when training dedicated models (ex: a classification model for sentiment analysis).
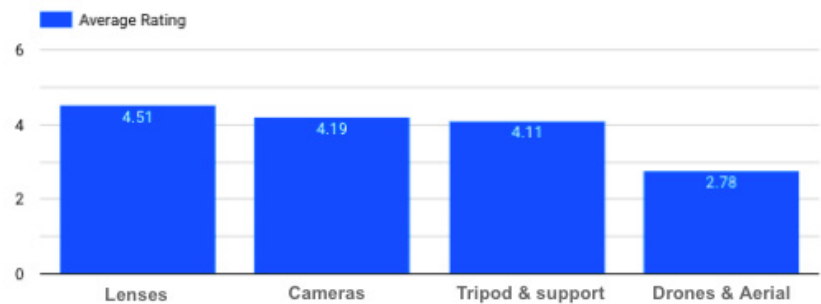
WHAT'S THE RATING DISTRIBUTION OF EACH CATEGORY?

We can see here that Lenses have the highest average rating, while there are a lot of negative reviews (especially with a score of 1) for Drones and Aerial Imaging.
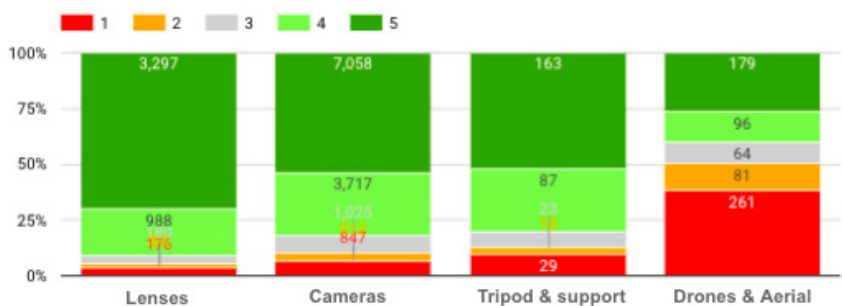
Average rating & distribution of each product category

Using NLP to understand your customers' concerns

Now, to understand what the reviews are about, we will implement the different NLP approaches mentioned previously.

DATA CLEANING

Before doing anything else, we need to clean the text data, to make it usable by the different NLP methods (this step is not always required, depending on the algorithms you want to use).

We applied standard pre-processing functions that were relevant to our data (removing HTML, punctuation, phone numbers, …), and we implemented a custom list of stop words that we remove from reviews (for instance the word "camera" does not bring that much information to our analysis).

You can find a lot of these functions in our NLPretext Github repository.

MINING INSIGHTS IN A FEW LINES OF CODE

Now that we have for each review:

• A product category

• The review original text

• The review cleaned text

• The review cleaned text split into tokens

• The product rating

We can start by simply looking at our most frequent words (single words, bi-grams, tri-grams…). It's a simple analysis, but it gives you an immediate vision of what the main topics are for each score and category.

```python
from collections import Counter
import matplotlib.pyplot as plt
import wordcloud

plt.rcParams["figure.figsize"] = [16, 9]


def create_ngrams(token_list, nb_elements):
    """
    Create n-grams for list of tokens
    Parameters
    ----------
    token_list : list
        list of strings
    nb_elements :
        number of elements in the n-gram
    Returns
    -------
    Generator
        generator of all n-grams
    """
    ngrams = zip(*[token_list[index_token:] for index_token in range(nb_
elements)])
    return (" ".join(ngram) for ngram in ngrams)


def frequent_words(list_words, ngrams_number=1, number_top_words=10):
    """
    Create n-grams for list of tokens
    Parameters
    ----------
    ngrams_number : int
    number_top_words : int
        output dataframe length
    Returns
    -------
    DataFrame
        Dataframe with the entities and their frequencies.
    """
    frequent = []
    if ngrams_number == 1:
        pass
    elif ngrams_number >= 2:
        list_words = create_ngrams(list_words, ngrams_number)
    else:
        raise ValueError("number of n-grams should be >= 1")
    counter = Counter(list_words)
    frequent = counter.most_common(number_top_words)
    return frequent


def make_word_cloud(text_or_counter, stop_words=None):
    if isinstance(text_or_counter, str):
        word_cloud = wordcloud.WordCloud(stopwords=stop_words).
generate(text_or_counter)
    else:
        if stop_words is not None:
            text_or_counter = Counter(word for word in text_or_counter if
word not in stop_words)
        word_cloud = wordcloud.WordCloud(stopwords=stop_words).generate_
from_frequencies(text_or_counter)
    plt.imshow(word_cloud)
    plt.axis("off")
    plt.show()
```

WordCloud

Leveraging these functions, we can easily display a Word Cloud of most frequent words, using reviews for Cameras with a score between 1 and 2:

Then display a similar Word Cloud using reviews for Cameras with a score between 4 and 5 :

We can easily identify the main points brought up in both cases.

For reviews with low scores, we have a lot of mentions about the battery, the device screen, its price or even mentions of a real bug encountered.

For reviews with high scores, we see that the photo quality, and the functionalities or design are being brought up often

We could do this exercise for each product our company has, in order to see the specificity of each and be able to draw conclusions at a more granular level.

N-grams Count

We can also use the frequent_words function to display the most frequent words, bi-grams or tri-grams:

To go further, you could then put in place a function displaying the reviews associated with a keyword, in order to zoom in on n-grams you find interesting. You could also look at n-grams with the highest / lowest TF-IDF (easy to compute with the sklearn library), since it allows you to see important words based on a different metric than a simple frequency counter.

Sentiment Analysis

Next, we move on to a sentiment analysis approach. Usually, it is used to predict if a text is positive or negative. In our case, we already have this information (the score between 1
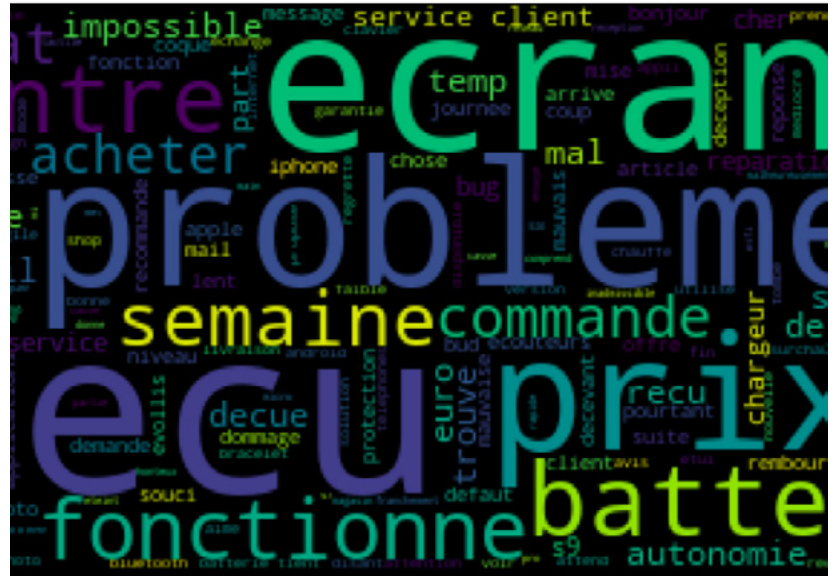
and 5 gives us the sentiment behind the review). But training a model to predict this rating will help us find which words (features) are key for customers.

What we can do is to train a sentiment analysis classifier on this data, and then use libraries like SHAP or LIME to understand which features (= words) have the most impact on a review being classified as positive or negative.

CLASSIFIER

To train a classifier, you have a lot of possible algorithms you can use, ranging from the classic sklearn LogisticRegression, to ULM-fit models (see this notebook to train a French ULM-fit model, and this article to understand more about ULM-fit) or the Ludwig classifier developed by Uber.

You might want to start with a simple one first, to see if it already answers your needs, before putting in place more complex algorithms.

Make sure to take into consideration the fact that your dataset is probably imbalanced (more positive than negative reviews, in our case).

FEATURE IMPORTANCE

Once your classifier is implemented, you can move on to the most important step: getting insights from features importance.

In the following example we apply SHAP on our model (here, a simple sklearn LogisticRegression):

We can see here that the functionalities, photo quality, and zoom features have a really positive impact on our clients' satisfaction, while the flash, memory card or batteries tend to have a really negative impact when mentioned in a review.

Words like "excellent", "perfect" or "bad" were removed from this analysis (before training the classifier), because they will be considered as the most important features, while in our case we want to focus on finding insights about our products, not really improve our classifier performance.

See this notebook for an example on how to use SHAP, with a public dataset.

BUSINESS THEMES IMPACT

Our third approach was kind of different from the previous ones, as it

starts from business-related themes chosen by someone knowledgeable when it comes to the products.

The point is to analyse how predefined business themes impact products ratings, to understand if they are a source of strength or an issue to solve.

DETERMINING THEMES

The first step is to classify the reviews into the thematic categories. Either by labelling your dataset manually (then you could train a classifier if you want to automatically classify new review into themes), or with a rule-based model.

In our case we used a rule-based model because it can already bring up good results at low cost (e.g: if you're curious about your lenses quality or your after-sales services, it can be simple to establish rules that will determine if a review mention those or not).

THEME IMPACT

In a second step you can compute your global average score, then the average score of reviews talking about a specific theme.

By subtracting both scores, you can deduce the impact your theme has on your global score.

We should here worry about our after-sales service because it is often mentioned in a negative way (though it could also be because people contacting the after-sales service often had an issue in the first place. Which is why, you should then look into detail at the reviews mentioning this theme, to really understand why it was brought up).

Here again, business knowledge is essential to make sense of your results.

On the other hand, when our designs or lenses are mentioned, it's often linked to a review with a high score, which could mean it's one of our strengths.

Use SHAP to explain features impact

```
X = vectorizer.transform(corpus)
explainer = shap.LinearExplainer(model, X, feature_perturbation="interventional")
shap_values = explainer.shap_values(X)
X_array = X.toarray()
```

```
shap.summary_plot(shap_values, X_array, feature_names=vectorizer.get_feature_names())
```
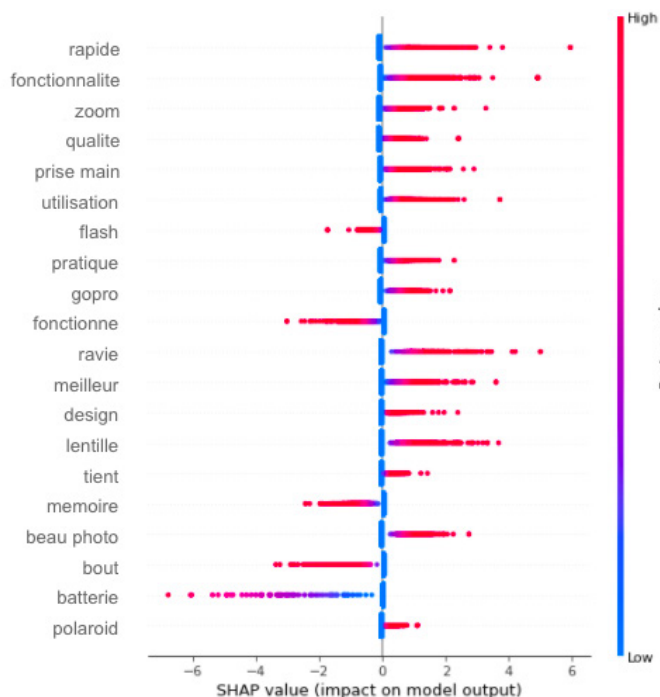
HomeServe

CASE STUDY

HOMESERVE

Using speech analytics to improve customer satisfaction

"The detection of non-compliance in sales calls use case analysis allowed us to prove that AI can be leveraged to better orient the work of the compliance team."

CHALLENGES

Present in France for 20 years, HomeServe is the world leader in home insurance services, with 8 million customers and over one billion in revenue.

When it comes to home emergencies, the most common channel used by customers is the phone – 9 out of 10 customers prefer it. This particularity places the call centre at the heart of every step of the insurance value chain, from sales to customer service, and ultimately assistance.

Although HomeServe has already developed AI-based conversational solutions and is present on Google Assistant and Amazon Alexa, they wanted to explore new ways in which AI could improve efficiency and customer experience in their existing phone channel.

They were especially interested to see what impact speech analytics could have on the vast amounts of unexploited customer data they collected.

SOLUTION

Artefact began by helping HomeServe opt for a "make" over "buy" strategy, as only a proprietary asset tailored to their organisation, combining technology and skills, could meet their many objectives, which include:

We also laid out a plan for developing HomeServe's expertise in natural language, data science algorithms, and AI data-treatment technical structures.

Next, Artefact set up a long-term multidisciplinary team with HomeServe comprised of a business team, a core data team, and an IT team to assess the maturity of speech analytics, the value and feasibility of relevant use cases, and improvements to customer experience and efficiency.

Because we couldn't build the entire architecture right away, we needed to quickly demonstrate the value of speech analytics to all stakeholders via a minimum viable product (MVP), able to expand after its validation with business experts.

To do this, we analysed two high-value use cases in a four-week cross-company workshop. We developed several microservices for data collection and processing and packaged to enable these use cases to be developed and be reused in the future, should the MVP phase prove successful.

1. **Refining understanding of customer contact root causes**
2. **Detecting risks of non-compliance within sales calls**

RESULTS

The most important conclusion for Artefact is that we proved the technology is mature. Speech analytics is ready to produce value for companies right now.

The customer contact root cause use case analysis produced three actionable insights, which could help call centre agents to perform better, sell more contracts, and benefit from a less tedious workload:

The detection of non-compliance in sales calls use case analysis allowed us to prove that AI can be leveraged to better orient the work of the compliance team.

AI Industry Solutions

Data for Finance & Industry

Unlocking the Future: How **financial institutions** can prepare to scale AI

Athena Sharma
Consulting Director & Global Financial Services Lead
ARTEFACT

According to The Economist, some 54% of large financial institutions (FIs) had already adopted artificial intelligence back in 2020, so imagine where those numbers stand today. To add to that proliferation, 86% of financial executives say that they plan on increasing AI investment through 2025. And in another survey, 81% said that unlocking value from AI would be the key differentiator between winners and losers in the banking industry.

"There's clearly a very strong value case to be made for AI in financial institutions", said Athena. "Investment banks are perhaps the earliest adopters and beneficiaries of machine learning technology in the algorithmic trading space. After all, 70% of FIs now use machine learning for fraud detection, credit scoring, or predicting cash flow events, and conversational AI is commonly used in retail banking and insurance. Yet despite this, many FIs fall short when it comes to productionising their AI projects to deliver concrete, enterprise-wide value."

Athena explained the main challenges to AI project success and how to overcome them:

• Number one requires investing in core technology and data management.

• Number two involves implementing a future-oriented operating model.

• Number three concerns proactively considering AI ethics and regulation.

Investing in core technology and data management

For Athena, one of the key difficulties FIs face is that their core technology is built for traditional operations, such as payments, lending, claims management. "Legacy IT stacks don't have the flexibility to deploy AI skills. The computational capacity for data management and analytics you need in a closed loop VR application just isn't there and testing and developing AI technologies can take days or even months – prohibitive when you're trying to be innovative. The solution?

Change core technologies: move to cloud computing.

"A cloud environment can reduce the time it takes to test and develop AI solutions down to a few minutes, thanks to managed services", assures Athena. "A bank I worked with started transitioning into the cloud two years ago, and their innovation rate has increased by about 49% according to their own KPIs. That might seem small, but for an incumbent, monolithic institution, it's quite revolutionary."

Another facet of this challenge is investing in data management – both in terms of data quality and data access. In FIs, data is siloed across various business units and divisions. As a result, data isn't standardised, quality is hard to manage, and there's no single source of truth, so stakeholders are unsure if the underlying data of proposed projects is trustworthy. "Investment in modern data governance and data management practices is crucial for FIs", insists Athena. "And a key component of that is what we call an Enterprise Data Model, or EDM. It's not an IT concept, but a way of describing and logically organising your data – all of your data – in business relevant language – a kind of business glossary, if you will, that streamlines data quality management for all certified users." The final part of this challenge is data access.

"Data is the most valuable raw material any organisation possesses; key to leveraging its value is to have access to analytics at scale, at the point of decision making. It's especially difficult in banks due to data confidentiality. An innovative solution is to create API-enabled databases for more effective and secure data access, but at scale and in real time to fulfil your business objectives, and in real time to fulfil your business objectives."

65

Implementing a future-oriented operating model

The second challenge for financial institutions lies in the operating model they use. Most are organised according to business divisions, often with centralised IT functions, impeding their ability to innovate. Business leaders set their own agendas and AI strategies, resulting in fragmented teams and a waterfall approach that leads to delays, cost overruns, suboptimal performance and a total lack of a test and learn mindset. FIs must be able to work in an iterative manner to continuously innovate and improve – a necessity in order to scale AI, because no one ever gets it right the first time.

"Instead, we at Artefact propose a more agile and flexible future-oriented operating model based on data products. A data product is essentially a set of data solutions that directly address a business challenge or business outcome. Each data product is developed by a dedicated team that has their own budget, assets, and KPIs."

"For example, say you have a client 360 team of business, IT and data stakeholders. They can provide several data products to the business, as well as to external customers, so you obtain a customer 360 analytics layer. Data scientists and engineers can use this analytics layer to test and learn AI ML solutions. You could also have a client 360 dashboard with relevant KPIs for your frontline sales colleagues and use it to improve customer lifetime value. You could also provide data to your marketing team about optimization and personalization to help them better spend their budgets."

The possibilities are endless, but in essence, a modular operating model allows your teams to better collaborate and work towards a common strategic goal, rather than in the silos that currently divide FIs – as well as a myriad of companies across all sectors where product teams are not yet a reality.

Proactively considering AI ethics and regulation

Investment in AI ethics and regulation is crucial for financial institutions right now. In reviewing the European Commission's proposed Artificial Intelligence Act, the European Data Protection Supervisor (EDPS) considers that stronger protection of fundamental rights is necessary, including strengthening the protection of individuals' fundamental rights, including the rights to privacy and to the protection of personal data.

Regulatory restrictions are to be imposed on anyone who uses any software associated with biometric technology in financial institutions, human capital management or credit assessment of individuals. As things stand, this will affect almost all FIs. While the full extent of future AI regulation is not yet clear to anyone, what is evident is that regulations will be ethics-based. But many leaders in the financial services industry feel their companies don't understand the ethical issues associated with AI.

Artefact proposes developing an ethical in-house AI governance framework that covers all aspects of AI ethics, including buyers, including data management, model training and retraining AI explainability. To do this, expert advice may be useful, but what's really needed is a two-part mindset shift covering all aspects of AI ethics, including buyers, data management, model training and retraining AI explainability.

The first shift requires large-scale stakeholder buy-in, by obliging stakeholders to let go of the siloed mentality, division and operating models that are preventing you from productionizing AI. The second is moving from a risk-averse to a pioneering mindset. This requires a deep cultural change where the entire organisation attains a high level of literacy on the impact of AI, its applications and its ethics, in order to be innovative without being irresponsible.

"It isn't easy, especially in an industry where risk aversion is deeply embedded. But ultimately, when it comes to AI adoption, I don't think financial institutions have much optionality, it's not how or if AI can add value to your business. It's about how you can embed AI in your day-to-day operations in order to remain relevant and competitive in a rapidly changing global marketplace."

Gaining buy-in for data & analytics initiatives in **financial services**
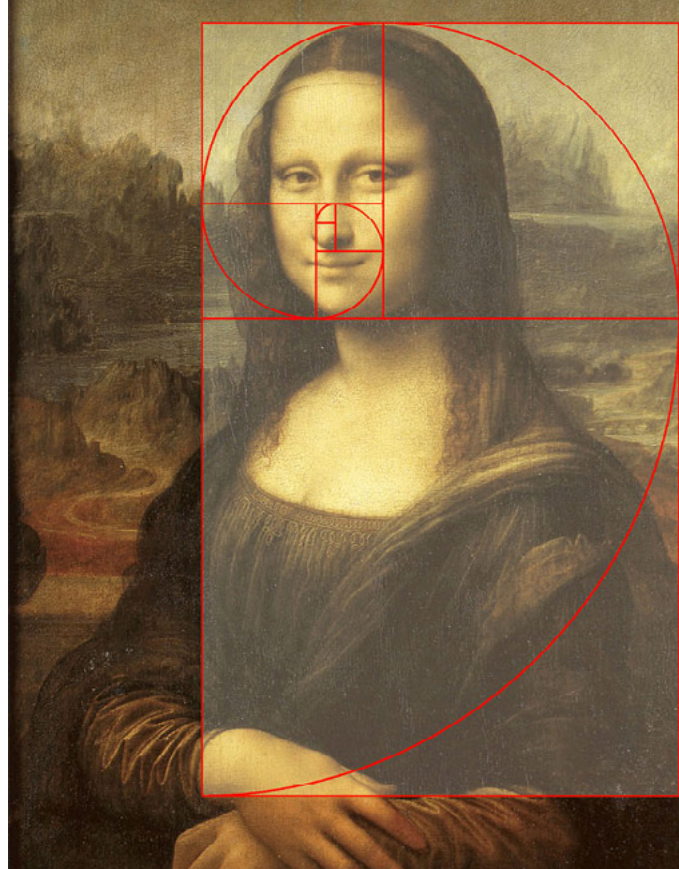
Akhilesh Kale
Director – US FSI Lead
ARTEFACT

Corentin Boinnot
Senior Consultant
ARTEFACT

John Ly
Senior Consultant
ARTEFACT

Modern day da(ta) Vincis

Picture Da Vinci's Mona Lisa. The world has been captivated by the inherent mystery tied to this work of art: her elusive smile and eyes that seem to follow you. What is less obvious is the science, math & geometry Da Vinci used to anchor this masterpiece. Today's Data Leaders are the modern day Da(ta) Vincis. They apply complex data/AI machinery behind the scenes to offer compelling business outcomes. This analogy underscores the complexities within Renaissance paintings and data ecosystems alike. When it comes to data, however, it is the business outcomes that matter most.

Setting the scene

ADVANCED ANALYTICS IN FINANCIAL SERVICES

Financial institutions were some of the earliest to invest in their data capabilities, driven by regulatory mandates and huge cost saving opportunities. Industry surveys continue to show investments in data and AI & the associated expectations are on the rise.

- 94% of companies are increasing their data investments in 2023 (*Data & Analytics leadership annual executive survey – New Vantage Partners – 2023*)

- There is an estimated US$447B cost saving opportunity for banks in 2023 (*Data For Finance – Artefact Research – 2022*)

- All banks, regardless of size, are widely adopting AI and Machine Learning (*The state of AI in 2022 – Mckinsey & Co – 2022*)

Strategic decision-making is a delicate dance of decisions driven by gut and by machines & data. It is paramount to anchor to fundamental value drivers before any conceptualizing and pitching take place.

This article outlines some practical keys to leverage the power of data & analytics to transform your organization. The objective is twofold: **you want to receive buy-in and harness maximum value from data projects.**

Speak the language of the business

The majority of data leaders have a technical background and possess a genuine passion for their area of expertise. Nonetheless, it is key to resist the urge to drop complex data concepts when pitching a data transformation proposal to business leadership. Business leaders do not always understand data jargon, and might be thrown off by a technical approach to solving their problems. They care about the business outcomes and impacts: increasing sales, reducing operating costs, freeing up human resources and mitigating risks. Be sure to speak their language!

EXAMPLE

If you are trying to secure buy-in to launch an enterprise-wide data transformation project:

– Do not: talk about microservice architectures, data observability, domain driven ownership or common data governance principles.

– Instead: reference the cross-selling opportunities, the time saved by teams cleaning data and the opportunity to reduce customer churn.

Bring business and technology onboard

Just as our Renaissance hero combined art and science, modern day Da(ta) Vincis need to recruit experts from across the business to instill a data-driven culture into an organization. Understanding the strengths of colleagues – and what motivates them – can help pave the way for effective collaboration, for example:

> **Business talent is needed to spearhead the vision and quantify the value impact:**

- Pique their interest with compelling use cases through a new data platform.

- Make it clear they would own the project roadmap.

> **Engineering & IT excel at building complex tools:**

- To ensure they are onboard and invested, offer them the opportunity to roll out the latest data & AI technology.

EXAMPLE

Let's consider a scenario where you are trying to build a next generation Customer Data Platform and need to attract talent from elsewhere in the business to drive the project:

- *Have a clear plan of who you need and what they can do.*

- *Understand what will drive and motivate individual experts to work on the project.*

- *Provide incentives to work with the data team, make sure everyone benefits.*

Business could be hooked by the use cases available through a new data platform. They would own the roadmap. IT would love to roll out the latest data & AI tech.

Advanced Analytics in service of your strategy

DEFENSE

Risk Monitoring & Mitigation

Regulatory

Compliance

Reduce Costs

OFFENSE

Increased Revenue

More Retention

Competitive

Advantage

Unite the ranks, mobilize the leaders

In addition to the right mix of talent, long term project success relies on support from leadership as well as collaboration from on-the-ground teams to drive the project forward:

• Receive buy-in from leadership. Highlight the key issues it can solve and bring them into a vision they can relate to.

• Identify data champions within the business to facilitate collaboration with your data teams. Their knowledge of both data and business will allow them to educate colleagues, clarify directives and identify relevant use cases.

• Establish a deep working relationship between data professionals and their business counterparts. Encourage the following key conversations between data and business teams:

> Help business teams understand the real cost of delivering current data insights.

> Have your data teams illustrate the benefits of their initiatives to the business to help ensure a smooth, self-driven onboarding process.

> Rely on pilots and proof-of-concepts today to demonstrate the feasibility of capabilities at-scale in the future.

BRINGING LEADERSHIP TEAMS ON BOARD FOR A DATA QUALITY PROGRAM

Get creative to launch new initiatives. Help leaders and their teams to understand current challenges and the feasibility of practical solutions.

SITUATION

Poor data quality | Ineffective data products | No buy-in from leadership

TOP-DOWN

CDO briefed c-suite on lost revenues and increasing costs due to poor DQ
····· Quantified lost revenue & high costs due to poor data quality

To demonstrate the time wasted, each leader made to clean a real world data set in excel

High costs of poor dq was understood | a practical solution was socialized
····· **3 DOMAINS PRIORITIZED FOR LAUNCHING THE DATA QUALITY PROGRAM**

Demonstrated impact to the business analysts through a do engine pilot

CDO team engaged business analysts to understand impacts caused due to poor DQ
····· Hybrid pods
Sprint retrospects
Ask me anything hours
Coffee-chats
Surveys, slack

BOTTOM-UP

Aim big, start small

LAYING THE GROUNDWORK FOR SUCCESS

The Mona Lisa wasn't painted in a day. Data Leaders won't be able to reorient their whole company to make it data driven overnight. Have your Da(ta) Vincis invest their energy and resources into first sketching out the story. Use this to build a high value use case that highlights all the advantages of leveraging your data insights.

No matter how flashy the use case, the transformation from sketch to masterpiece won't take if you don't provide the business with the means to get invested in data. They will need:

• Clean data or the means to clean it

• Business user friendly tools

• Data team's time, attention and resources

ARTEFACT CLIENT ILLUSTRATION

How one useful Dashboard convinced our client to go 'all in' on data visualization

Discovering analytics gold

To complement a strong case – or generate buy-in from the business without one – identifying existing pots of "analytics gold" in your organization is a good approach. There are almost certainly passionate individuals in the business who have developed impactful analytics solutions locally. Find them and industrialize their initiatives so the whole company can benefit.
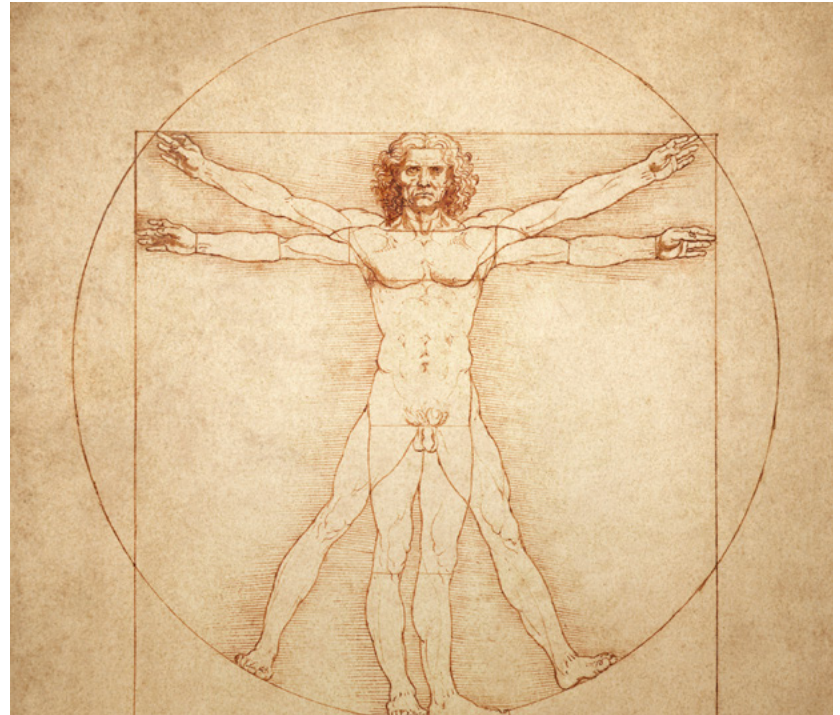
Our client team (Commercial Sales Operations) spent 5 hours a week manually building their weekly B2B Sales Report

Using a lean data visualization POD team, we built and deployed an automated B2B Sales Power BI Dashboard in three weeks leveraging the existing data eco-system with no disruption

Our client **reorganized their sales operations** to leverage data from this dashboard

Marketing now wanted a dashboard to visualize omni-channel campaign results on a unified interface

We helped the client **build a BI center of excellence** to backlog, prioritize and execute on enterprise needs
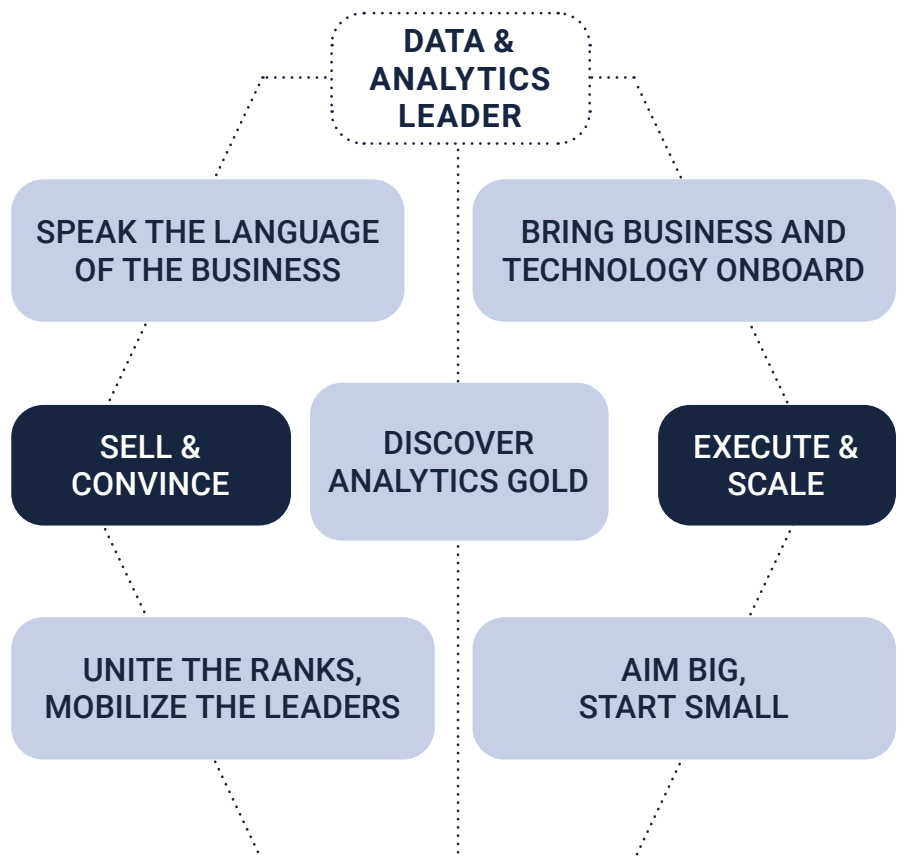
| FIND | Organize business hackathons |
| | Organize slack/teams channels for ideas sharing |

| ENCOURAGE | Publish success stories in newsletters, bring teams to leadership townhalls |
| | Push program branding |

| REWARD | Get leadership to appreciate |
| | Extend to performance reviews and internal incentives |

Here are a few tips to help unearth and identify these hidden treasures:

• Build an enticing platform (newsletter, monthly data meetup, …) so people will want to share their ideas.

• Establish recognition and rewards to better acknowledge their ideas and work.

• Work with HR to ensure data practices are measured in managerial performance reviews.

• Build guilds and organize dedicated chat channels to share knowledge.

Finding pots of (analytics) gold hidden in the business

Data & analytics leaders, true renaissance thinkers

Being a remarkable "sales" person is a quality that makes one stand out, no matter the profession. However, being remarkable at "selling" is less about the act of sales and more about the ability to solve complex problems. The role of a data leader encompasses all of these traits. They need to be an outstanding salesperson, a smooth operator and possess the ability to understand their organization's numerous challenges.

Going back to where we started this blog, we see the analytics leaders today as modern day Da(ta) Vincis. Powered by the tools of the 21st century – data, analytics and AI, these true Renaissance thinkers are creatively painting the business impact story through the complex machinery of indisputable facts.

DATA & ANALYTICS LEADER

SPEAK THE LANGUAGE OF THE BUSINESS

BRING BUSINESS AND TECHNOLOGY ONBOARD

SELL & CONVINCE

DISCOVER ANALYTICS GOLD

EXECUTE & SCALE

UNITE THE RANKS, MOBILIZE THE LEADERS

AIM BIG, START SMALL

The road ahead: data-driven sales is critical for the evolving car industry

Axel Tasciyan
Data Consulting Director
Automotive Lead
ARTEFACT

This requires a 3-pronged approach: sell more, sell better, seek out new sales streams.

The commercial model for selling cars is shifting, with manufacturers adopting the direct-to-consumer (also know as the agency model) trend being witnessed by a wide variety of business sectors as people look to simplify the buying process.

This is a sea change for vehicle producers. It's also complicated to navigate in terms of the agreements that will still be required with dealerships.

The net result is that automakers are experiencing major and sustained pressure on their bottom lines.

Tackling the industry's shifting vehicle retail sands requires a three-pronged approach: sell more; sell better; seek out new sales streams.

Of course, it looks obvious on paper. But far from being a rousing sales team pep talk, these concrete steps form the basis of the modern motor trade. Let's look at them in detail.

Sell more

This is not the obvious admonishment to sell more vehicles (although that of course is important); rather it focuses on increasing the 'extras' sold to someone who has purchased a car.

Currently a buyer visiting the manufacturer's website a week after placing their order is likely to be offered another car. This doesn't work for anyone.

Contrast that with a consumer buying a car who, on returning to the website,

more to already acquired customers and realise more profit is a key piece of the most effective digital strategy.

Sell better

Currently carmakers control only a small part of the purchasing experience.

A potential buyer might approach the manufacturer, only to be guided to the nearest car dealership, where they test drive the car, negotiate the price, agree a sale, provide their details and with whom they arrange subsequent services. This information is all vital — but it belongs to the distributor.

However, as DTC selling becomes more prevalent, data ownership will evolve and change, and manufacturers will preside over increasing amounts of valuable insight about their customers and prospects.

Armed with this information, the manufacturer can improve their sales conversion rate because they know more about who buys their cars and why — and therefore how to attract them and retain their interest throughout the sales funnel. At the same time, this knowledge sharpens their targeting ability and enables them to decrease their acquisition costs.

Ultimately "selling better" is about marketing, and within that truly understanding how to use digital tracking tools (such as Google analytics) in order to mine the wealth of information that already exists. Armed with these details, digital ad spend can be optimised based on real and accurate business and customer data (such as customer lifetime value, actual purchase details and online behaviours), rather than generic media KPIs.

Toyota Canada is one manufacturer actively demonstrating an advanced approach. Using a scoring model based on first party data such as online

is shown a range of available ancillary options — extended warranties, accessories, service specials etc (and even specific offerings such as seasonal items, based on their geographic location — snow-tires in winter for example).

This personalized approach, undertaken in close collaboration with dealers, extends the carmaker's relationship with the buyer from the current period of one to two years, potentially to between five and eight years. And by increasing this customer lifetime value, it opens up a range of revenue-generating opportunities that add value to the end user.

Looking outside the automotive sector provides further inspiration on how to 'sell more'.

Telecom company Orange, for example, sells a wide variety of accessories such as earbuds and phone cases as a result of the personalization of its website and after sales advertising campaigns.

And travel specialist Pierre & Vacances Center Parcs, operating a voucher system as the pandemic forced people to cancel their holidays, encouraged travelers to supplement trips planned for the future with incentives such as bigger rooms and additional activities.

Making these as appealing as possible required the company to minutely analyse its customer data in order to offer experiences that were relevant to different customers; leveraging the initial 'big' purchase to position the follow-up messages in order to sell

behavior, it identified the people with the highest propensity to buy a vehicle; targeting these new audiences drove a conversation rate six times higher than the previous tactic of re-engaging website visitors, while reducing the cost-per-acquisition by 80 percent. More advanced activity will include using increasing amounts of first party data to steer advertising campaigns.

Seek out new sales streams

Extending the car-buying experience gives manufacturers the opportunity to build up a picture of their customers — the environment they live in (rural or urban), the type of car they need, how they use it and potentially where they go.

Anonymizing these details and adding them to the vast amounts of first-party data that the industry already owns, puts manufacturers in charge of a lucrative, mainly unrealised, revenue stream that can be unlocked via data partnerships.

Data partnerships enable an organization to access the first-party data of another organization, either paying for it directly, or reciprocating with its own first-party data. The strategy enables both enterprises to explore new routes to new customers. For the automotive industry, partnerships with insurance companies are an obvious link, but this approach is easily broadened — hotels and holiday companies are key contenders for understanding more about peoples' driving and travelling habits for example.

The structure and organization of data partnerships is still new. It's therefore a value-generating differentiator for manufacturers that are prepared to be trailblazers and invest time and effort getting it right now before it reaches maturity.

In short, finding more value is about manufacturers complementing the traditional way of selling cars and

using creative thinking to make money from everything they know about the people that buy their products, in a way that enhances the customer experience.

The data 'glue'

So what is the link that holds together the modus operandi outlined above? Data. Data, data and more data.

This calls for a change of mindset in the car industry. Instead of one organization selling new cars, another selling used ones, another offering spare parts and yet another selling services, with each keeping their data separate, the vision of tomorrow relies on one central data hub that benefits everyone. Sometimes referred to as a Customer Data Platform (CDP), today's cloud solutions for the different tech components keep costs manageable, and make it feasible and relatively uncomplex for most enterprises.

With a CDP established, data management and analysis can deliver relevant insight that adds value and generates revenue opportunities throughout the complete customer lifetime.

However, this isn't a "tick box" exercise; the work of a data-marketer is never done... Data, and especially business data, is continually added to the CDP, and with it the picture of customers becomes clearer, richer and more granular — giving manufacturers and dealers ever more accurate tools with which to sell more, sell smart and develop new sales streams.

INTERVIEW

How Nissan is transforming in the digital world

In this Q+A, Dév Rishi Sahani, Nissan's Global Head of Customer Experience Data Analytics & Reporting, chats with Pascal Coggia, Artefact Partner and UK Managing Director, to explain how the Japanese car giant has accelerated its digital transformation over the last few years and is now using data and BI Hubs to drive operational efficiencies and sales around the world.

Pascal Coggia
CEO
ARTEFACT UK

Pascal Coggia: What does digital transformation mean for Nissan?

Dév Rishi Sahani: The term digital transformation is overused to an extent, so it depends on how you define it. At Nissan, we embarked on our digital transformation journey several years ago. We've done well on the technology transformation, and we will continue to do that, but the most important thing for us is keeping up with customers. Our entire focus is on what we call customer experience transformation, and if you look at it from that perspective then the challenges are very clear.

The first challenge concerns the organisational or operating structure. Thinking about customer journeys as a linear path is a traditional way to visualise how your customers interact with your brand, but, ultimately, it limits itself to be very transactional and locks you into silos. To move from transactions to relationships with customers, we must stop looking at the step conversions or handoffs between these channels and connect those silos.
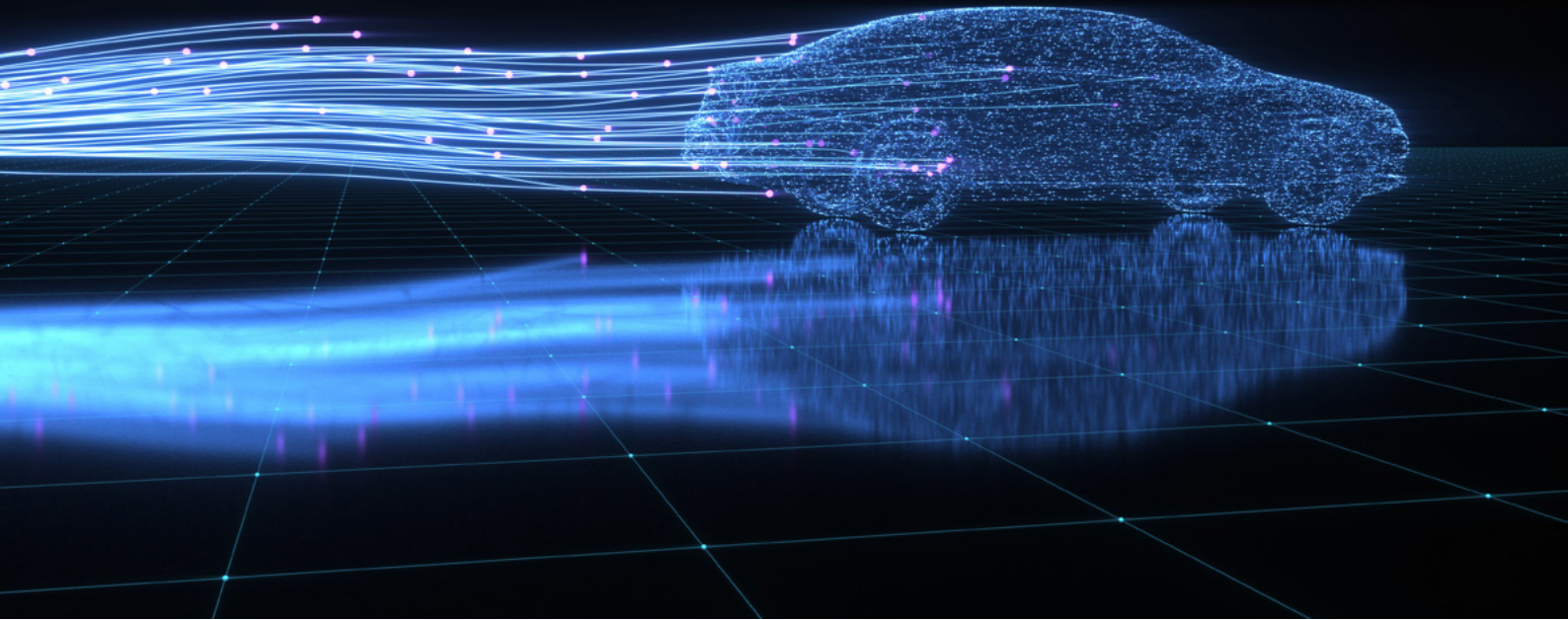
Pascal: Ok, can you tell us how you are using data? How are you becoming customer-centric?

Dév: Over my career, I hold a firm belief that the amount of data available will always be ahead of organisational appetite to consume and action those insights. This is not a bad or a good thing, it's just reality. Sometimes the battle is simply because the data conflicts with the intended course of action you wanted to take.

But more importantly, it's when analysis does not present clear implications or recommendations — if the data is not telling you what to do next, it's pretty useless. This is where the word 'utilisation' becomes relevant; it shifts the focus from how many people are using the data to the utility. We've been leading the adoption of data within our organisation by keeping a clear definition of its usefulness for our markets, business functions and digital teams.

At one end is the input data, like dashboards and the support systems that enable upstream planning and decision making; on the other end is output data, which are the predictive models and the data science that shows the operational results of our initiatives. What links the two of these together is a data-driven, hypothesis-led test-and-learn culture.

Pascal: Can you give us an example of a data project that has resonated at scale within your organisation?

Dév: In one way it's been our entire journey for the last two or three years... developing CEDAR, our internal brand for the data analytics function. It stands for Customer Experience Data Analytics and Reporting. It helps us turn data into information, knowledge, and wisdom. That's where the actionability and usability come in.

For the first couple of years, we were kind of hardcore, getting everything organised. The first 'Eureka!' moment was when the whole organisation realised they could look at the data and see consumer trends across 147 markets. Rather than make assumptions based on samples of 5,000 people, say, we can now see how one million visitors interact, every single minute of every single day.

A couple of years ago, we launched a new car in a specific market and spotted a lot of cross-segment buying going on. We could see when customers changed their minds from the model we thought they would buy to the one they actually did. We could also determine who would buy an automatic transmission or a manual transmission, for example. We can now consume data and see significant patterns in the consumer buying journey. We want to know what our customers want and respond to that.

Pascal: Can you tell me more about this CEDAR dashboard — this BI hub? Why does it exist?

Dév: It wasn't the first thing we built. We started by focussing on data – what questions can we ask the data? what answers will the data give? We got our head around that in the first year of the program. Then it was about 'democratising' this data. We realised that having the data was not the challenge, but making sense of it

CEDAR HOME

was. This is where we developed our partnership with Artefact to create CEDAR home.

CEDAR is an independent platform that works across Nissan, and across functions, and contains lots of different dashboards and insights. It's our hub of experience and consumer insights – where our people go to see the world as our customers do. Anyone in the company can go there and if you don't have a login, you just sign up. It should feel just as easy as browsing our website.
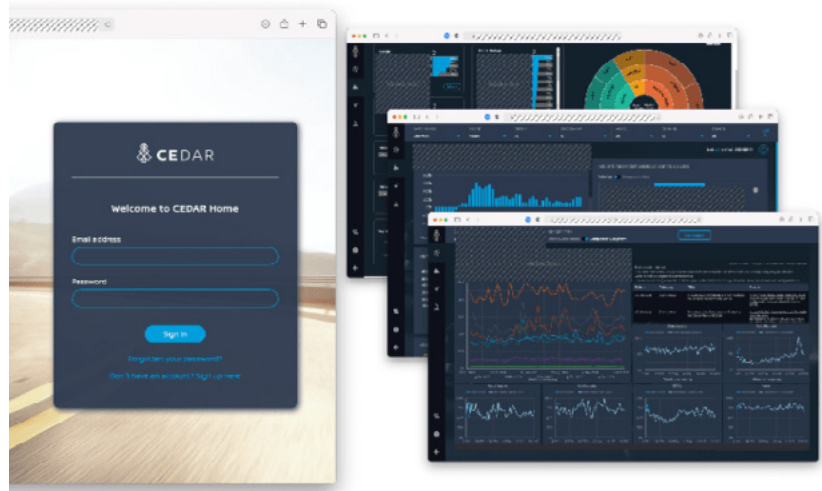
Most importantly, as a data person, we can look at the analytics in CEDAR to understand how it is being used, where the points of friction are, where people get lost, and what parts are useful or not useful. Rather than just ask, we know.

Pascal: Have you seen people using CEDAR to make decisions they were unable to before? Or is it just helping them save time?

Dév: We have a lot of people who are making good decisions, but the evidence of people using CEDAR to make different decisions or decisions that they could not take before is quite satisfying.

We started this journey from a customer experience focus perspective. We wanted anyone who was working on customer experience to use these insights. But very quickly it has spread to teams from other functions such as advanced product planning and market intelligence. All these different functions, at global and market levels, know about CEDAR. That itself is giving us a lot of context about how data is becoming more useful across our organisation and is being used by different types of business functions.

Pascal: Tell me a little bit about the team. Who's in the driving seat?

Dév: Data is not everything; you need to know how to unlock insights from it. This is where having the right team helps. When I joined Nissan in November 2017, there were two people in the global data practice and from there we grew the team across three core pillars — called Measure, Optimise and Predict.

The Measure pillar is tasked with understanding how we organise the data and turn it into information. We grew up in a very collaborative way with partnerships, like with Artefact and other companies. We knew we needed the machine to build that up. The second pillar is called Optimise and is responsible for running data-driven experiments and testing and learning – which we do a lot of. The third pillar is called Predict, which is the data science pillar. We were very careful to add that at the end to ensure we didn't rush ahead and bring in talent that would get very bored, very quickly, because we didn't have the right use cases lined up.

Now, the direct team is mostly based in the global customer experience centre in London, along with some folks in our headquarters in Yokohama, Japan, and we work with embedded analytics teams across all markets.

Pascal: So, what does success look like? How do you see your data projects progressing over the next couple of years?

Dév: Success is about making the data we generate useful for both Nissan and its customers. That involves tying everything we do back to our organisational KPIs. We can attribute where we are starting to make a difference from an organisational perspective, and, equally, we can tie it back to customer quality metrics that show how we get short-term wins or long-term longevity. That's what success of the customer experience program looks like. The success of the data program is about continuing to make it very simple and keeping it growing.

ARTEFACT

Data for Impact

Use data to measure and reduce your environmental impact with Artefact

Léonard Cahon
Senior Consulting Manager
ARTEFACT

It's challenging for companies to consistently measure their environmental impact and prioritize the actions they need to take to reduce it. But with data and artificial intelligence, it's possible to accurately measure the impact of an activity and effectively guide decision making while adopting best practices for increased energy sobriety in digital business. This is what Artefact, a leading international data transformation and consulting company, demonstrates.

Companies often struggle to measure their environmental impact. It may be difficult for them to gather all the data they need, or they may simply not have control over it. They can't know, for example, how their customers will use their products once they've purchased them and what impact this will have on the environment. To better estimate this impact, companies can turn to Artefact's data experts. They advise large companies on how to turn data into business value and focus more on the potential of data to positively impact the environment.

An increasingly restrictive legislative framework

By 2024, companies will be required to produce and publish their non-financial information, including their annual carbon footprint.

"The legislative framework is increasingly restrictive, but it sets objectives and constraints without a common frame of reference for the same sectors to achieve these objectives and respect these constraints," explains Margot Millory, Consulting Manager in charge of Sustainability at Artefact.

Two companies might end up with extremely different orders of magnitude simply because they don't have the same measurement methods or don't take the same emission factors into account. Artefact offers to assist companies by capitalizing on the entire data value chain, from data strategy to data governance to implementation, all with real consulting support to guide decision making towards greater efficiency. The company is already working with many sectors (consumer products, luxury goods, telecommunications, etc.), but more and more business activities will be impacted by these issues, as evidenced by the CSRD (Corporate Sustainability Reporting Directive), which now sets the standards for non-financial reporting for 50,000 companies in Europe.

Data governance for energy sobriety in digital activity

As with many of the challenges facing companies, the creation of a reliable, sustainable database is an essential prerequisite for implementing a strategy to reduce environmental impact. Artefact's sustainable data governance offer enables its clients to benefit from a clean and structured data repository, a real added value for the organization.

"This data is often spread across several databases that are rarely or never consulted, with no clearly established governance. To retrieve them on a regular basis and track them over time, companies need strong data governance skills, which is the hallmark of a business like Artefact's. You can't measure if you don't have a solid, sustainable data structure." explains Vincent Luciani, CEO of Artefact.

This exercise is essential for companies to identify missing data and get up to speed in order to collect and structure it and improve its quality. As a result, they'll be able to develop automated and sustainable reporting tools to serve as the foundation for building a realistic environmental impact reduction trajectory.

Best practices for responsible digital activity

Artefact's clients are increasingly integrating extra-financial logic into their reasoning. In response to this evolution, the company also assists them in their efforts to reduce their carbon footprint through use cases based on data science and AI: inventory optimization, waste reduction and improved delivery times.

Digital technology is currently responsible for 2.5% of national GHG (greenhouse gas) emissions, and these could increase by 60% by 2040. Artefact wants to minimize the negative external factors linked to its activity through the promotion of best practices to implement data/AI projects responsibly, in a logic of energy sobriety. To this end, Vincent Luciani and Margot Millory have joined the Institut Numérique Responsable (Responsible Digital Institute), a think tank that focuses, among other things, on reducing the economic, social and environmental footprint of digital technology.

"It's crucial to operate in both an energy-efficient and ethical way, whether in terms of data collection, computing power or the way we build and deploy our algorithms" explains Vincent Luciani.

Environmental initiatives at the heart of future work models

Today, Google, Microsoft and Amazon are all capable of completely transforming the market with new technologies. Artefact is optimistic about these changes, convinced that technology can offer many benefits, and that those who produce it are equally aware of environmental issues.

"We believe that the environmental transformation wave will be as big as the digital wave." explains Vincent Luciani.

It's likely that companies will launch environmental initiatives or place the environment at the heart of their work model. Artefact sees this as a motivating factor: a new way of thinking with different constraints, different stakeholders, and multiple priorities, as different KPIs are involved. The companies supported by Artefact used to have a customer-centric approach, which puts the customer at the heart of the company's activity. But their priorities are changing, driven by growing pressure from consumers who now demand full transparency and concrete action from the companies whose goods and services they consume. "Our business is evolving towards an approach that puts the citizen and accountability at the center of the company's activity," explains Vincent Luciani.

Education: a pillar of sustainable digital transformation

Companies are seeking to recruit trained people to work on data-related projects; there is also a real need for training in sustainable development and environmental impact. Through its Artefact School of Data, which teaches data-related jobs to people in professional retraining, Artefact tackles important notions such as adopting best practices in order to implement frugal algorithms, measuring the carbon impact of an e-commerce chain or establishing an accurate sales forecast in order to prevent food waste, as carried out in the bakery and pastry department of Carrefour hypermarkets.

Industrializing carbon footprint measurement to achieve **neutrality**

To achieve carbon neutrality, the challenge for large companies is first to track their carbon footprints. Some large companies have initiated a change in the culture of data processing to achieve the industrialization of this data, which is massive, heterogeneous and rarely prioritized.

Vincent Blaclard
Partner
ARTEFACT

The climate emergency has become a major issue for our society. Recent events, in particular the multiple shortages and repeated heat waves, only confirm the acceleration of current and future difficulties that must be overcome. Today, many European companies listed on the stock exchange are announcing their commitment to climate transition. 30% have made a real commitment to reduce their carbon emissions, but it is estimated that only 5% of them are on track to do so. It is not a simple exercise. Reducing emissions in a sustainable way requires accurate measurement of their carbon footprint, in order to develop concrete actions. At Artefact, we believe that exploiting the data to its full potential is a major asset for the success of this approach.

Achieving carbon neutrality with three objectives thanks to data

Let's take the example of the Carrefour group, for whom we are carrying out an assignment. Carrefour's ambition is to become the world leader in food transition, particularly in e-commerce. One of its major objectives is to make e-commerce carbon-neutral by 2030. Three main levers of action have been identified in order to reach these objectives: reducing Carrefour's own emissions, engaging its service providers to reduce their emissions and finally encouraging its customers to adopt eco-responsible behaviors. This ambition, in addition to responding to the climate emergency, also has a strong economic impact.

We must meet the expectations of consumers, who are increasingly committed, and anticipate the tightening of the legislative framework to come, such as the eco-score that will become mandatory from 2023 for certain players. In order to face these challenges, Carrefour understood that it was necessary to have a measure of the carbon footprint: to make a quantified inventory of the starting point, to determine the impact of reduction initiatives and to be able to communicate both internally and externally on the successes, and also on the challenges to come. This measure will be the compass for the 2030 neutrality trajectory. It will have to meet the requirements of reliability and transparency, and allow for the implementation of concrete actions.

The major challenge of prioritizing data

A large part of the project's efforts consisted of collecting a large amount of very heterogeneous data from multiple sources (for example, mileage data of delivery services or IT infrastructure emissions data), in order to orchestrate them and build a consolidated carbon footprint measurement. The goal is to obtain a comprehensive measurement of all emission items for each individual order. The main difficulty with any project of this type is the complexity of accessing data that can be used quickly. Most large groups have already launched significant programs to better govern the data, addressing quality and accessibility issues first. These programs are often very large and obviously cannot handle all the data created in a company, often very large. Prioritization of data domains closest to the core business is necessary, such as sales, supplier or consumer data.

"Reducing your emissions in a sustainable and lasting way requires you to accurately measure your carbon footprint."

Unfortunately, data related to sustainable development is rarely prioritized in such initiatives, as it is rarely used in an industrial way by large groups. Today, a team of experts needs several weeks of project time to calculate a carbon footprint measurement that is often static. It is certain that tomorrow all companies will have to be able to calculate this carbon footprint at any time, in the same way that companies are required to be financially transparent.

The parallel with the data market

We can take the parallel further with the evolution of the data market. Ten years ago, awareness of data in large companies was still limited. It was the exclusive territory of small teams within the IT or digital departments who worked on use cases, without the capacity to bring their solution to scale. Today, the importance of data is heard at the executive committee level of large groups, and is perceived as a strategic priority at all levels. This evolution has been, over the last ten years, the result of a collective awareness of the importance of data, notably through geopolitical and strategic issues, as well as tensions between major powers and large technology groups. This awareness has gradually taken hold in all organizations, even those less advanced in digital technology. It has been accelerated by the arrival of new generations (millennials) in decision-making positions, who have been aware of digital issues since their childhood.

Measuring the carbon footprint of all activities

This evolution is not going smoothly, and the use of data does not always give the expected results, often because robust foundations have not been put in place. The major groups now understood the importance of this fundamental work and are launching numerous programs on the subject.

We are undoubtedly at a crossroads as far as the ecological transition in companies is concerned. The successive disasters of the summer of 2022 are helping to accelerate this awareness, while a new generation of workers who are highly aware of these issues is entering the job market. Nearly 76% of Gen Yers place CSR above salary in their job search criteria and 70% are willing to pay up to 35% more for a sustainable, low-carbon product or service.

The market is still at this stage: there is a strong will to move forward, but the foundations needed to achieve these goals in a sustainable way often have to be built, which Carrefour has understood well. It is therefore crucial that companies equip themselves with the capabilities and tools to match their ambition, in particular the measurement of the carbon footprint of all their activities. This measurement must be industrialized, calculated in real time, accessible and integrated into all business processes. For example, the carbon footprint could be integrated into budgets and used to assess the impact of new projects, along with the revenues generated and the associated CapEx and OpEx costs.

Consolidating data governance

Once these foundations are built and consolidated, large corporations will be able to leverage their data much better to accelerate their green transition. Strong data foundations are a major prerequisite for deploying AI solutions at scale; the same is true for the green transition, where AI will certainly play a role once these foundations are consolidated. It's often more appealing to talk about AI than data governance, but I am convinced that the success of these initiatives lies in the ability to move forward on both fronts: delivering impact through targeted initiatives, while building the right foundations to sustain those impacts.

Applying machine learning algorithms to satellite imagery for agriculture applications

Paul Devienne
Senior Data Scientist
ARTEFACT

This article will:

- Show you various applications of machine learning and computer vision to satellite images for agriculture.
- Present a series of algorithms to successfully detect and label agricultural plots.
- Suggest alternative methods depending on the availability of data.

This article assumes basic fundamentals in data science and computer vision.

Business motivation

A solution able to automatically detect and label crops can have a wide range of business applications. Computing the number of plots, their average size, the density of vegetation, the total surface area of specific crops, and plenty more indicators could serve various purposes. For example, public organizations could use these metrics for national statistics, while private farming companies could use them to estimate their potential market with a great level of detail.

Naturally, Satellite imagery was considered and identified as a very viable data source for 3 specific reasons:

Scalability: A bank of images covering the whole world is available right away and being updated regularly

Data richness: Satellite images can provide a lot more information than simple pictures. Instead of a 3-band image of Red, Green and Blue pixels, some satellites can provide more than 15 features per pixel

Cost: Even though satellite imagery can be quite costly, some options are fully free, such as Sentinel 2, which we ended up selecting as our main data source.

Step 1 — Detecting agricultural areas on satellite images

After retrieving and preprocessing Sentinel 2 images, our first challenge was to locate the plots and limit ourselves to specific areas of interest. Each image having a very high resolution, it would not be realistic to apply the whole processing to full size images. Instead, the first step to solve our problem was to crop large images into smaller fragments, and identify the areas where the plots were located on these smaller images. (cf image 2)

Solution 1A

TRAINING A PIXEL CLASSIFIER

The first solution for detecting agricultural zones on large images is to build a pixel classifier. For each pixel, this machine learning model would predict whether this pixel belongs to a forest, a city, water, a farm … and therefore, to an agricultural zone or not.

Because a lot of resources can be found for Sentinel-2, we were able to find labeled images with over 10 different classes of ground truth (forest, water, tundra, …). However, if the climate of your area of study is different from the area you trained your model on, you might have to reevaluate the classes attributed to each pixel.

For example, after training a model on temperate climate countries, and applying them to more arid regions of the world, we observed that what the model was seeing as forests and tundras were in fact agricultural crops.

Once your pixels are classified, you can drop all images that don't contain any agricultural areas.

Solution 1A pros:

• Most reliable and granular results (pixels)

Solution 1A cons:

• A dataset of labelled pixels is required

• Classifying each pixel generates a high computational cost

• Out of all available methods to detect agricultural zones, this one was the most accurate. However, if you do not have access to labeled images, we have identified two alternative solutions.
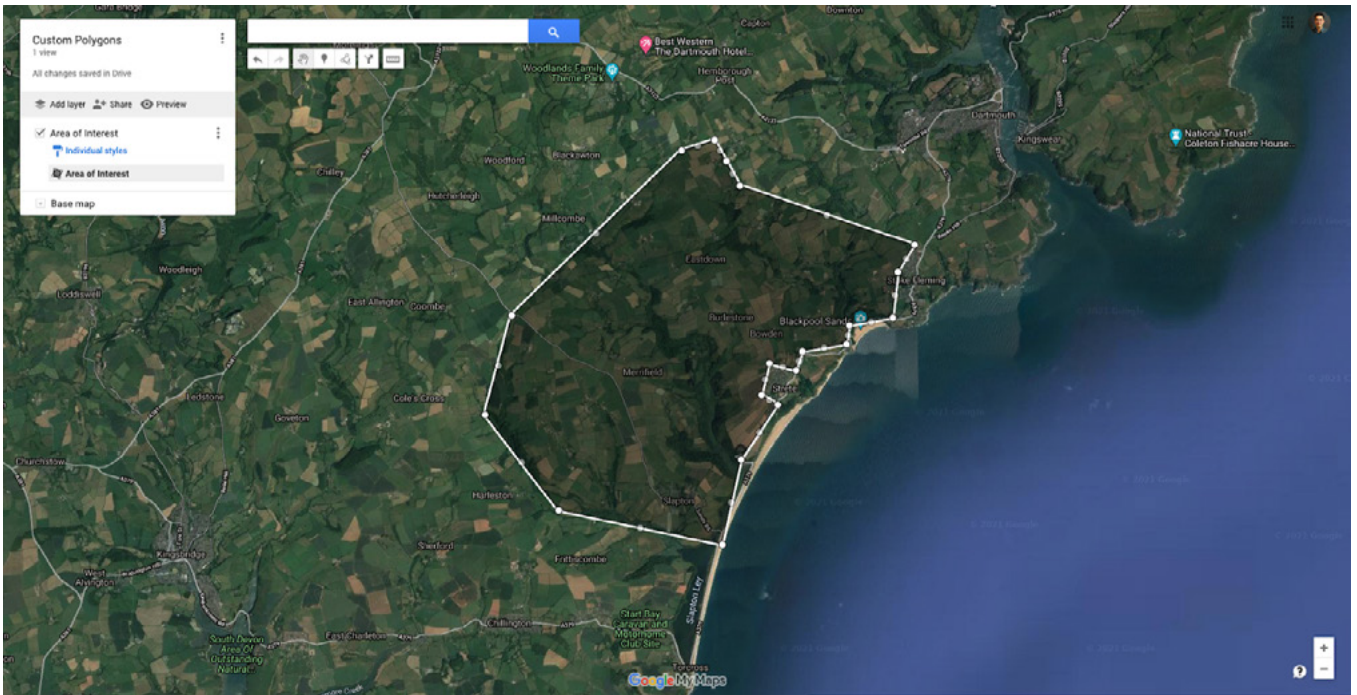

Image 2: Our desired output: fragments containing only agricultural areas (Copernicus Sentinel data 2019)

Illustration of pixel classification with 3 visible classes of pixels (Copernicus Sentinel data 2019)

You can design your own polygons on GoogleMaps, thus focusing on a specific area of choice while drawing around obstacles (water, cities …)

Solution 1B

MAPPING GEO COORDINATES TO PIXEL COORDINATES

If coordinates about your zone of interest have been labeled, or if you're labeling coordinates by yourself, it is possible to map these geo coordinates (latitude and longitude) to your images.

For example, if you have the coordinates associated with large farming areas, or if you draw large polygons on Google Maps yourself, you can easily obtain geo coordinates of agricultural areas. Then, all there

is to do is map those coordinates to your satellite images and filter your images to only cover the zones within your polygons.

Solution 1B pros:

• Also a reliable method

Solution 1B cons:

• You need a list of coordinates associated with agricultural regions

• Manually creating those coordinates can be time consuming

85
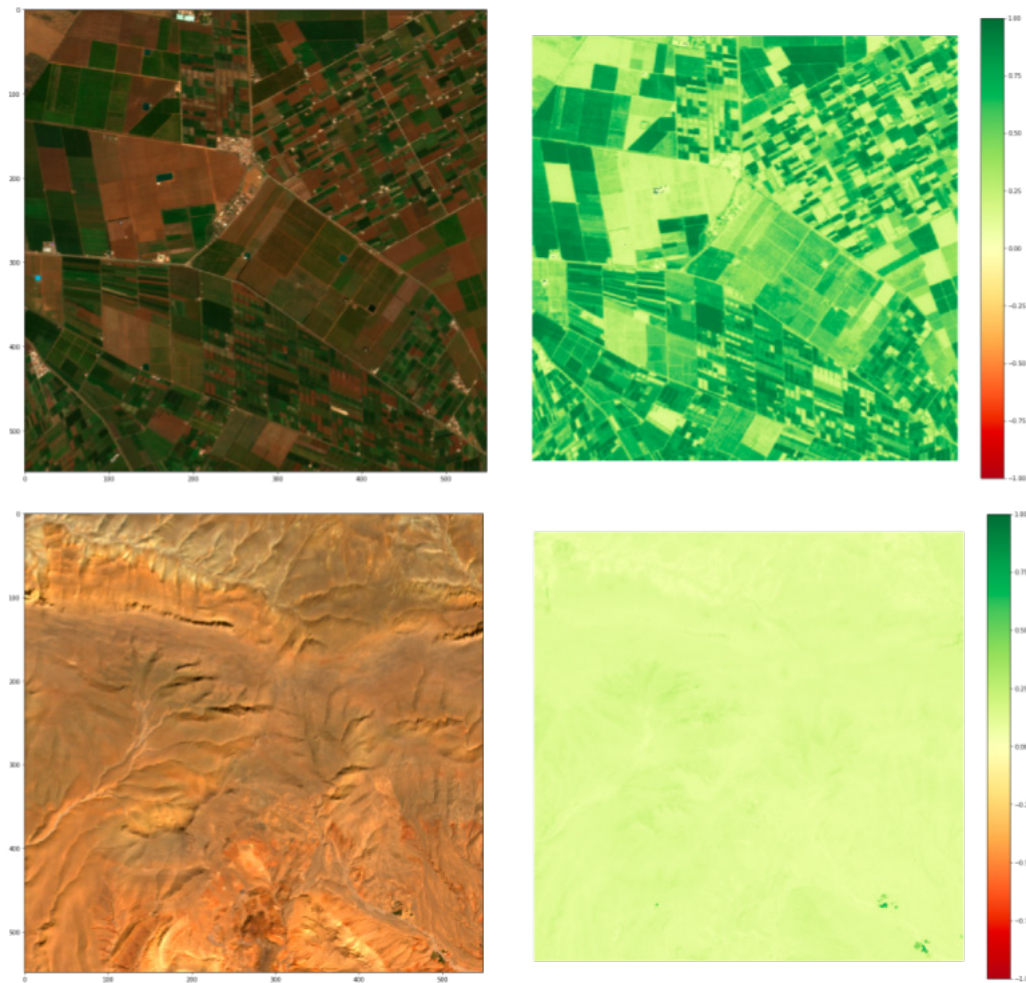
Visual representation of the NDVI on an agricultural zone and a desert
(Copernicus Sentinel data 2019)

Solution 1C

USING A VEGETATION INDEX

It is possible to compute a vegetation index from the color bands provided by the satellite images. A vegetation index is a formula combining multiple color bands, often highly correlated with the presence or density of vegetation (or other indicators such as the presence of water).

Multiple indices exist, but one of the most commonly used ones in an agricultural context is the NDVI (Normalized Difference Vegetation Index). This index is used to estimate the density of vegetation on the ground, which could serve to detect agricultural areas over a large image.

After computing NDVI values for each pixel, you can set a threshold to quickly eliminate pixels with no vegetation. We used NDVI as an example, but experimenting with various indices could help achieve better results.

Note that computing a vegetation index can provide you with useful information to enrich your analysis, even if you have already implemented another way to detect agricultural areas.
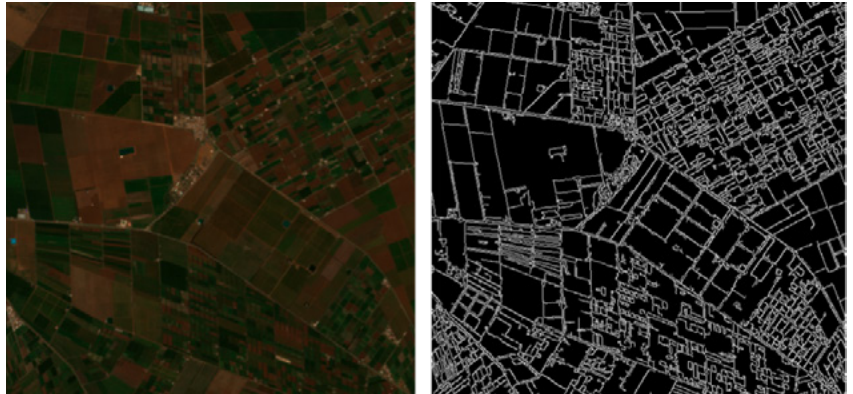
Solution 1C pros:

- Absolutely no labelled data required

Solution 1C cons:

- Not very accurate: for example, it could be hard to differentiate agricultural crops from forests

- The thresholds have to be fine tuned depending on climate and other specificities

An example of edge detection on agricultural plots using OpenCV (Copernicus Sentinel data 2019)

Step 2 — Detecting and outlining agricultural plots

BUILDING AN UNSUPERVISED EDGE DETECTOR

Once you have determined the location of your agricultural zones, you can start focusing on outlining individual plots on these specific areas.

In the absence of labeled data, we decided to go for an unsupervised approach based on OpenCV's Canny Edge detection. Edge detection consists in looking at a specific pixel and comparing it to the ones around it. If the contrast with neighboring pixels is high, then the pixel can be considered as an edge.

Once all the pixels that could potentially be true edges have been identified, we can start smoothing out the edges and try to form polygons. As expected, the performance of the edge detection algorithm is proven to be much better when applied to large plots:

This method allowed us to automatically identify close to 7 000 plots in our area of interest. Because we used the pixel classification method (see step 1A), we were able to to separate real farm plots from other polygons, thus only retaining relevant data.
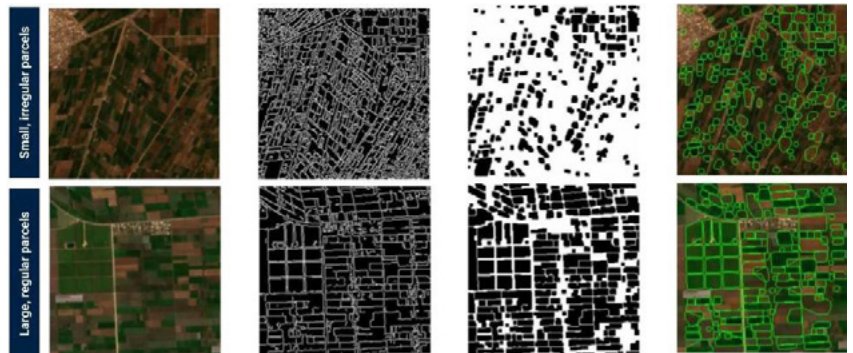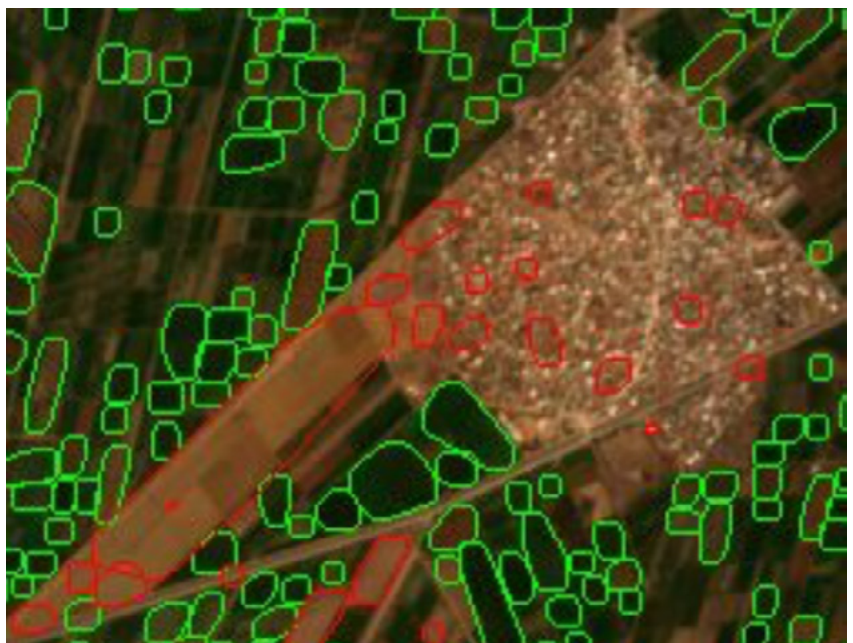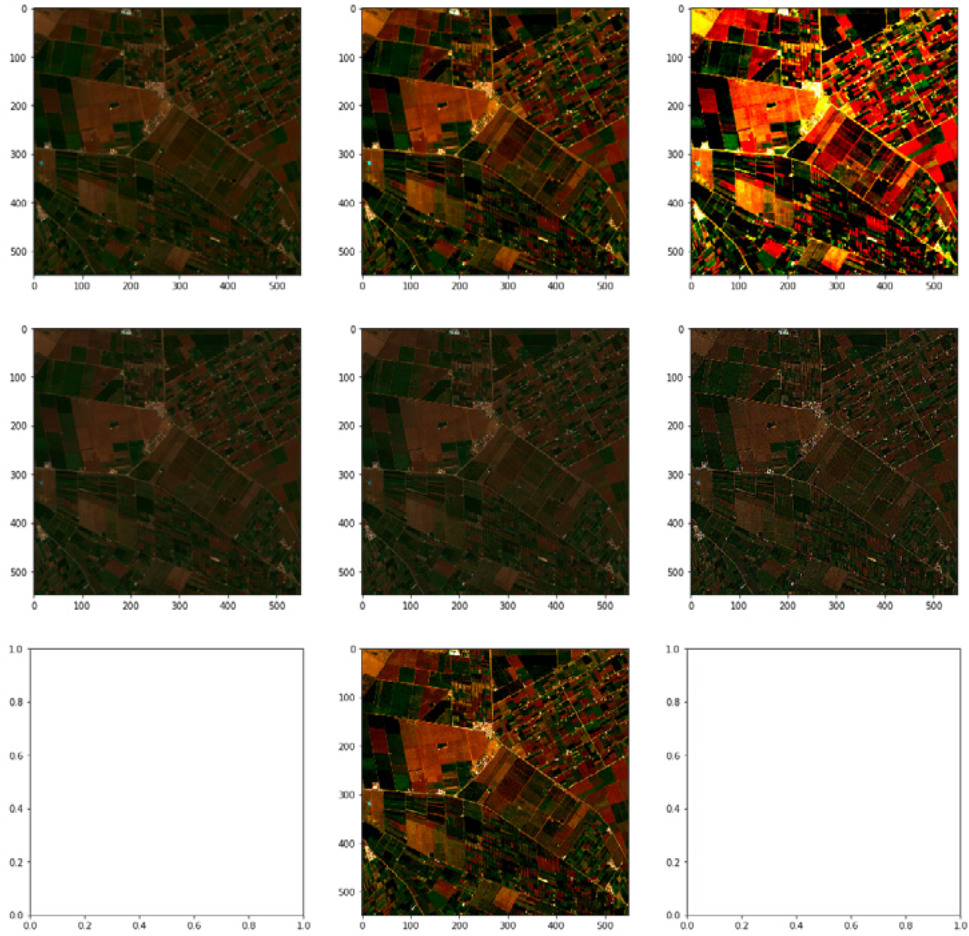

Illustration of the full process of outlining plots (Copernicus Sentinel data 2019)

Polygons consisting of a minority of "farm pixels" were eliminated (Copernicus Sentinel data 2019)
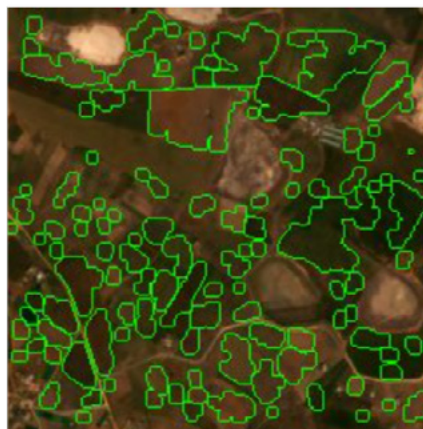
Experimenting on contrast, saturation or sharpness can help improve the efficiency of the edge detection (Copernicus Sentinel data 2019)

OPTIMIZING OF THE PERFORMANCE OF THE EDGE DETECTION ALGORITHM

In order to have the best possible results, it could prove useful to apply modifications to your image, notably by playing around with contrast, saturation or sharpness.

Another critical success factor is forcing the polygons to be convex. Most plots following regular shapes, forcing convex polygons can usually yield much better results.

Forcing convex shapes fits most plots much better (Copernicus Sentinel data 2019)

Latitude	Latitude	Latitude
44.864977	44.864977	44.864977
44.866378	44.866378	44.866378
8663378	8663378	8663378

Illustration of the external crop data source

Step 3 — Classifying each parcel to detect specific crops

Once all plots have been identified, you can now crop each of them and save them as individual image files. The next step is to train a classification model in order to distinguish each parcel based on its crop. In other words, trying to identify tomato crops from cereals, or potatoes.

BUILDING A LABELLED TRAINING SET

Because we did not have an already labelled dataset available, and because manually labelling hundreds of images would be too time consuming, we looked for complementary datasets containing the information about crops for specific plots at a given time and place.
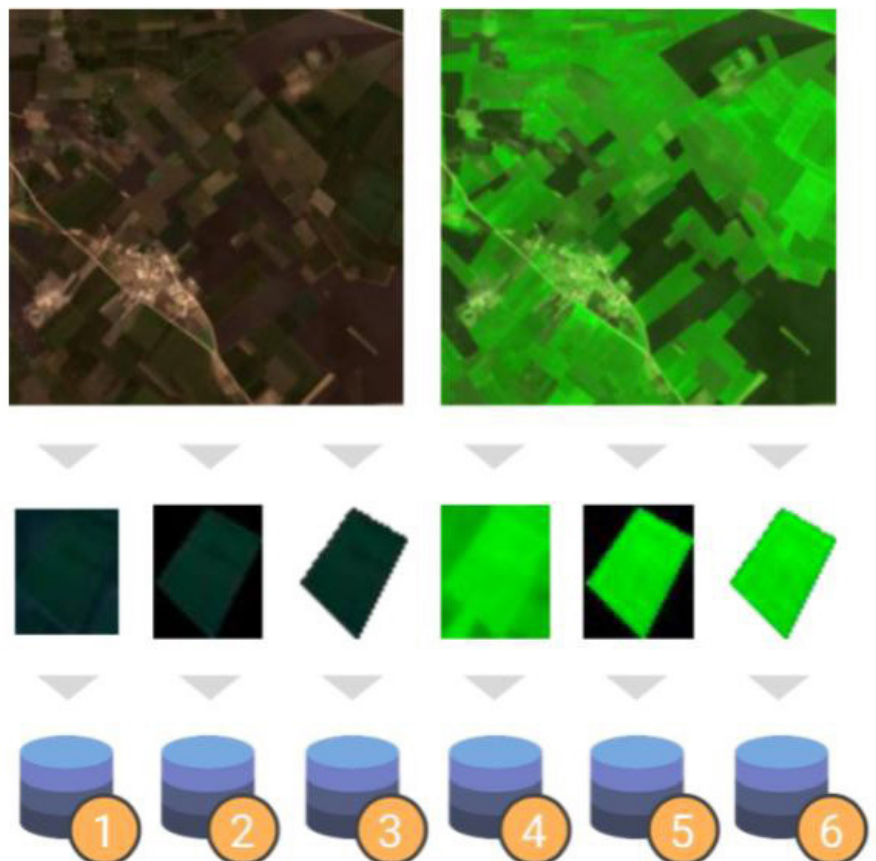
The ideal scenario would be to have pre-labelled images, but in our case we only had the geo coordinates and crops of a few hundred farm plots in our area of interest. This dataset contained a list of plots, the latitude and longitude of its centre, and the crop planted on it at a specific time of the year.

In order to build our training set, we used our geo coordinates to pixel coordinates converter to identify the specific plots for which we had a label (the crop) in our image bank.

Out of the 7 000 plots identified in Step 2, we managed to label around 500 plots thanks to our external data source. These 500 labelled plots served to train and evaluate the classification model.

Dozens of models were trained on datasets generated with various of data preparation techniques

When working with farm plots, just a few weeks can make a large difference (Copernicus Sentinel data 2019)

MODELIZATION

We chose to use a convolutional neural network using the fastai library, as it was an efficient way to classify our images.

In order to find the best possible classifier, we experimented with the input data:

- Selecting various combinations of color bands (Red, Green, Blue, Near infraRed …)

- Handling neighboring pixels in different ways: making them transparent, white, black … or leaving them untouched

After experimenting with various classification models, we reached 78% accuracy and 74% recall when performing binary classification on the smallest plots (and thus the hardest to classify due to the low number of pixels).

CHALLENGES TO KEEP IN MIND

When working with farm plots, even a few weeks can make a substantial difference. Within a few weeks, wheat crops can go from green to gold to harvested:

Thus, there are two things to keep in mind in order to replicate this project throughout the year:

- You have to build a model for each period of the year.

- Your labelled data containing information about the crops need to be refreshed regularly.

Conclusion

Working with satellite images opens up an endless range of possibilities. Considering how each Satellite provides different features, and how the availability and format of complementary data can vary throughout the world depending on your area of study, every single project will end up as a unique use case.

ARTEFACT
AI IS ABOUT PEOPLE

WE OFFER END-TO-END DATA & AI SERVICES

DATA READINESS

- ∧ Data Strategy & Organization
- ∧ Data Governance
- ∧ Data Platform
- ∧ Data for Sustainability
- ∧ Data & AI Adoption
- ∧ The Artefact School of Data

AI ACCELERATION

- ∧ AI Strategy
- ∧ AI & Analytics Factory
- ∧ Generative AI Factory
- ∧ AI for Operations
- ∧ AI for Customer Care
- ∧ A Supply Chain 2.0

DATA & DIGITAL MARKETING

- ∧ CX & Creative Studio
- ∧ Customer 360 & CDP
- ∧ MROI & Measurement
- ∧ Digital Analytics
- ∧ Media Buying & Certifications
- ∧ Retail Media

TECHNOLOGIES

- ∧ The Artefact Research Center
- ∧ Skaff platform: Tech Accelerators
- ∧ AI Partners
- ∧ Cloud Certifications
- ∧ Data & AI technologies (Fact)
- ∧ MLOps

FMCG • RETAIL & ECOMMERCE • LUXURY & COSMETICS • SPORTS & ENTERTAINMENT • TRAVEL & TOURISM • PUBLIC & GOVERNMENT • REAL ESTATE • MANUFACTURING & UTILITIES
HEALTHCARE • BANKING & INSURANCE • TELECOMMUNICATIONS

ARTEFACT

AI IS ABOUT PEOPLE

CONTACT

hello@artefact.com
artefact.com/contact-us

ARTEFACT HEADQUARTERS

19, rue Richer
75009 — Paris
France

artefact.com