# Technical Specification Double-side Page

1. TECHNICAL SCOPE:

'IMPROVING STORES' EFFICIENCY USING CLIENTS SHOPPING TIMES' challenge aims to identify improvement opportunities both at the commercial and operational levels. We have been solving these problems within the industry by computing the 'time in store' per customer according to their shopping basket. First, our proposed solution provides the Retailer with a methodology + Dashboards which give different perspectives to give a full understanding of retail stores and different buyers. Using this methodology, the first step would be to identify MIRROR STORES with a clustering analysis made among their different Retail branches (Hiper, Super, Proximity).

This will help to identify the stores with similar patterns (customer behaviour and operations of the store), to help us plan and execute different experiments. This way, the impact could be measured after applying the analytics + actions. Therefore the impact of the solutions provided can be measured with A/B testing with mirror vs target stores in an automated data flow. With this layer of BI Dashboard together with the data experiments + strategy, we will suggest different experiments that we have applied successfully in the past with other clients based on our Auto-ML algorithms. We will try to actuate in two different levels: Operational improvement (reduce time in-store), Sales improvement (increase sales)

As a value-added feature to the proposed solution, our promotional Reward-Engine (Roiward.io) can be linked to their SIGA platform to send different rewards to customers to improve shopping performance (sales and time). As well we can import the feedback of the reward engine campaign data to provide original customer data, giving the ability to the end-user to analyse the effect of different campaigns of the retail consumers.
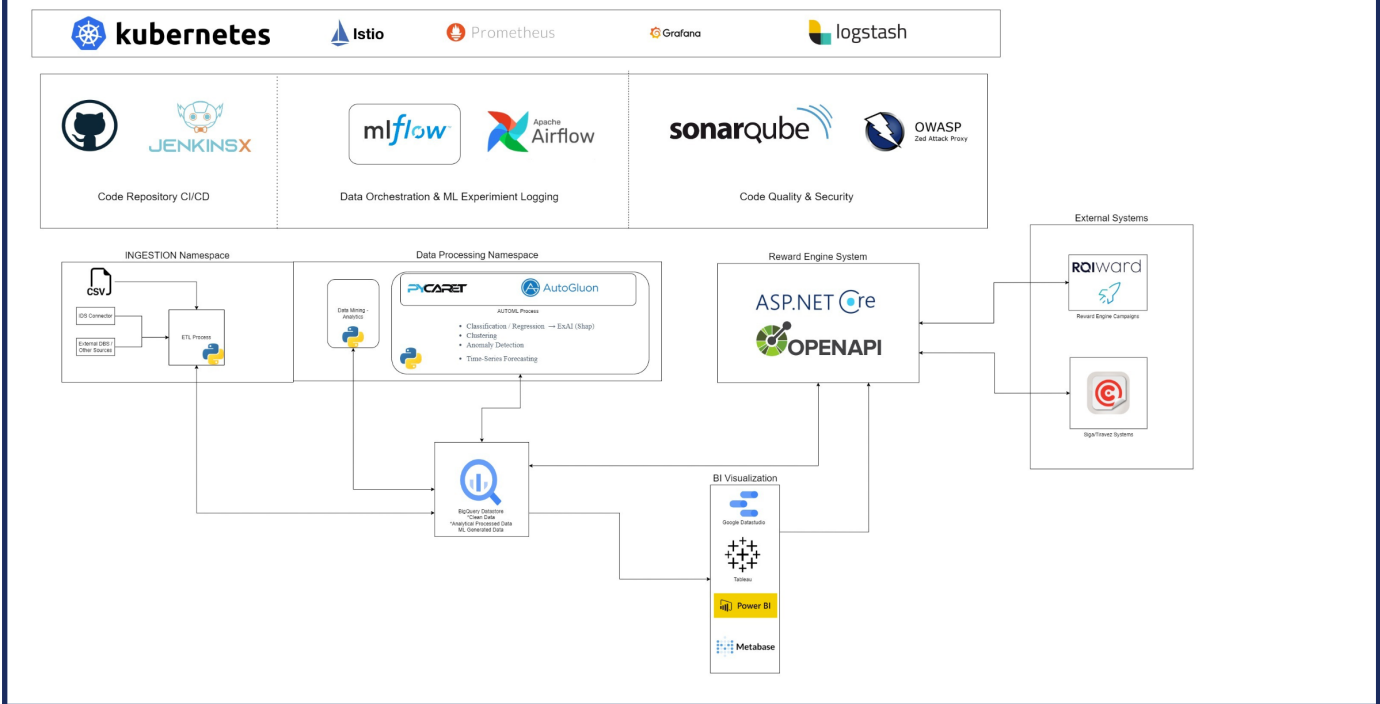
2. ALGORITHMS, TOOLS AND CONCLUSIONS:

As discussed with our data provider, our solution has two different areas of focus: Operational efficiency and Sales improvement. We plan to achieve these goals by facing the challenge of covering these different areas of growth with a proposal that lies first on layer Data and experimentation Strategy & Methodology. Here we define a field of data processing flows that set the base of the experiments and metrics in order to set up different growth activities. Once we have defined this flow we will plan and execute the following

ANALYSIS AND EXPERIMENTS PLANNED:

1) BI Reporting Data Flow in order to set the bases to experiment and report automatically the KPIs of the tracked metrics with the experiments

2) MirrorStores Analysis: AutoML Flow + AutoExplainable AI (SHAP) in order to identify the best retailers (Mirror Stores)

3) Customer Clusters (same analysis as above but oriented to Product selection + Time in Store + Sales) instead of retail performance.

4) In-store/ In-app promos: Market basket promotions within their SIGA App. Algorithm: MLXTend - FPGrowth. More frequently occurring items will have better chances of selling. With this, we can induce promotions on the customers within their Purchasing process

5) New Customer dynamic. EcoTempo: Speed checkout within their SIGA App with gamification of the rewards based on the accuracy of their usage of their app. Together with our proprietary REWARD API, we intend to reduce up to 15min of shopping time for each customer while maintaining low fraud rates.

6) Category management: Assortment and Inventory Optimization: From Predictive Choice Models to Near-Optimal Algorithms

3. SCALABILITY AND FLEXIBILITY OF THE SOLUTION:

The solution consists of a series of totally integrated microservices, designed to work on docker containers under a Kubernetes infrastructure, and using any BI tool (Google Datastudio, PowerBI, Metabase, etc). In the data storage aspect, Google Bigquery has been selected as the data storage system due to its price and scalability, although other solutions could be applied depending on technical needs. This architecture can adapt to almost any scale in the processing and data storage areas, using different auto scaling techniques to consume just the needed resources. For the solution to work with different data sources and shapes, there's a need to normalise the input data using ETL pipelines, which can be analysed during the EXPERIMENTATION phase.

Even if the project's first phases are aimed to give specific business outcomes in the physical retail sector, the usage of multiple AutoML libraries will allow the solution to provide different insights and valuable data working with different ML models easily, giving the ability to test dozens of ML models with the same data to find the one that adapts the best. This functionality is set to be defined clearly in the 'EXPERIMENT' phase. This could lead to a more reliable solution for other domains outside of physical retail. (Online retail stores, warehouse optimization, as examples). On top of this, the usage of ExAI technologies (as SHAP), will better explain the ML models, with more clear and understandable results.

BI dashboards will be hand-made but with an automated flow behind them, so that when the 'experiment' phase is over, there will be no need for manual effort in the end. In the first phases of the project, a more robust system with built-in multi-tenancy capabilities for data and services, and automatic dashboard generation could be possible if aligned with business and/or technical needs. If during the 'EXPERIMENT' phase we see that BI tool dashboards are not enough, a custom ReactJS frontend can be developed to achieve business objectives.

4. DATA GOVERNANCE AND LEGAL COMPLIANCE:

The proposed solution does not require any external exposures. Thus, available security infrastructure and policies within the mentioned architecture components will be enough for security assurance on data governance. However, within our ISO27001 certification process, we established internal processes, policies, and rules for user authentication along with monitoring layers that ensure the data governance of the solution complies with legal requirements that could appear during the project (these layers will be explained further in detail). Although we will not have to end-user personal data, regarding Legal Compliance and specifically GDPR, private and sensitive data feeding the solution is managed by SONAE.MC itself, and its treatment consent should be handled by them barring their legitimate interests.

From Dative.io we'll need to establish a contract with them which must include the appointment for data processing. We already work with personal information with other customers, and thus everyone within Dative.io is committed to confidentiality. Our organisation is ISO standard certifies and all our employees have authorization and control measures in place. In addition, it is important to notice that within the provided dataset there is no demographic information or any kind of customer details which could raise ethical issues. However, we will be vigilant in that respect and will react with moral responsibility if necessary, and despite all measures, we do cover ourselves with an insurance policy for up to 1.5M€ in damages due to data breach.

The data governance and legal compliance tools and systems we have in place are listed below:
- ISO 27001 Certification
- IDS: The International Data Spaces Association aims at open, federated data ecosystems and marketplaces ensuring data sovereignty for the creator of the data.
- Istio.io - Simplify observability, traffic management, security, and policy with the leading service mesh.
- Monitoring: Prometheus & Grafana
- Authentication: Google IAM System
- BigQuery - Accesses controlled by roles, rows and tables

5. QUALITY ASSURANCE AND RISK MANAGEMENT:

Dative.io team has more than 10 years of experience in the design, development, and implementation of data-driven marketing solutions with top tier Global customers e.g. P&G for **Retail projects** in Mercadona, ElCorteIngles, DIA, Carrefour... **Travel sector:** Lastminute.com, **Telco**: Euskaltel, Liberty Global International... Customers that have the highest quality and security standards. On the path of excellence, we certified our procedures and policies to improve and test our solutions and offer high-level security and quality assurance with ISO 27001.

In addition, Dative.io is composed of more than 15 international professionals with more than 5 nationalities on the team. As well we covered certifications like Google Cloud Architect and we have our own Cybersecurity team so we can take care of the projects starting with the best design principles and tools for monitoring them in real-time.

To minimise the risk and ensure that quality assurance and performance are upheld. We manage our solutions with different levels of software and systems that help us deliver safe and correct services:

- Security & access: Google IAM Auth system with 2FA and server log monitoring
- MLFlow in order to store, control, manage, update and measure the quality of our ML processes
- OWASP ZAP: Vulnerability Assessment, Penetration Testing, Runtime Testing, Code Review
- SonarQube is a Code Quality Assurance tool that collects and analyzes source code, and provides reports for the code quality of your project. It combines static and dynamic analysis tools and enables quality to be measured continually over time.
- Static code analysis within Github code and Docker images

On the other hand, as a referral and reference, part of this technology has been developed by an R&D project financed by EU FEDER Funds under the CDTI Cervera program in collaboration with Tecnalia Research and Innovation.