Master’s Degree thesis in

Engineering and Management

Politecnico di Torino

Development and Scaling of a smart control system to multiple products and multiple manufacture sites

Candidate **Maria Teresa Manente**

Academic Year 2023/2024, July 2023

Prof. Roberto Fontana	Supervisor - Politecnico di Torino
Dr. Gueorgui Mihaylov	Supervisor - Haleon UK

Acknowledgements

Sono grata al prof. Fontana che mi ha permesso di vivere questa fantastica esperienza londinese che porterò nel cuore per tutta la vita e, per la sua costante presenza sostenendomi in ogni fase del mio percorso. Un ringraziamento speciale va anche al dott. Mihaylov per avermi guidato con grande dedizione e professionalità fornendomi un punto di vista esperto che porterò con cura e attenzione per tutta la mia carriera. Con immenso affetto, ringrazio tutto il team di Data Science di Haleon UK, ed in particolare Christopher Fenton per la sua straordinaria competenza ed aiuto costante fornendomi un punto di riferimento quotidiano fondamentale per il successo di questo lavoro.

Voglio ringraziare di cuore la mia famiglia per il loro amore e sostegno incondizionato. La vostra forza e instancabile determinazione sono state una fonte costante di ispirazione e motivazione durante questi anni. Grazie per tutto il vostro supporto e per essere sempre al mio fianco.

Ed un grazie poi va Giovanni, che è stato la mia spalla e il mio rifugio costante. Sei stato al mio fianco senza mai mancare, giorno dopo giorno. Grazie per la tua presenza costante, per i tuoi saggi consigli e per aver condiviso con me ogni sfida e successo.

Declaration

It is stated that, due to a confidentiality agreement with the company, all sensitive data related to the batch production processes and information regarding the equipment of production sites have been intentionally omitted. This paper elaborates in detail on the logic behind each methodology conducted, by proceeding with the substitution of specific nomenclature with a general one. This approach aims to ensure clarity and comprehensibility of the content while adhering to the confidentiality constraints imposed by the agreement with the company.

Maria Teresa Manente
July 2024

Abstract

This thesis outlines the key aspects of developing and scaling a smart control system to integrate multiple products and manufacture sites based on an analysis performed in collaboration with the Data Science Department of Haleon UK. Nowadays consumers are looking for a broader variety of oral healthcare and in particular toothpastes. This trend towards products increasingly aligned with their needs contributes to an ever-increasing demand curve. The manufacturing processes of toothpastes are highly complex, involving key steps such as ingredient addition, mixing, control of pressure, temperature, and speeds. To proactively monitor these manufacturing processes, Haleon, a multinational consumer health care company, developed a nearly real-time smart control system that guarantees an optimal output not only in terms of product specifications but also in terms of process time, costs and resources. In today's business world, digital solutions play a crucial role in improving the operational efficiency and facilitating analysis and more sophisticated data management. This thesis project is representative of a standard real-world industrial problem of rapidly building and scaling a digital solution while accelerating adoption and delivery value. Effectively managing a broad and diversified production requires the control system to be easily and quickly scalable to integrate multiple products and manufacture sites. The thesis aims to identify the most relevant considerations faced during the development and scaling process of a control system from both technical and management standpoint. From a technical point of view, the thesis exposes not only the core of the data science model, which is the Fast Dynamic Time Warping, a machine learning algorithm that allows the proactive detection of steps of different processes and production sites, but it also provides an overview of the digital architecture that must be robust and well-designed to ensure that the different bottlenecks of the manufacture sites and requirements of all processes involved are taken into account. On the other hand, from a management perspective, given the high specificity and stringent quality standards of the manufacturing processes involved, the solution is developed in-house. The main management challenge lies in developing and simultaneously implementing an advanced digital solution with the aim of achieving accelerated

value generation. This challenge is facilitated by taking into account the needs of all stakeholders and by the accurate and informed decisions made throughout the project lifecycle (from the Proof of Concept up to the scaling phase) with a focus on the minimum viable produce (MVP), a simplified version of the final solution already containing all the key functionalities which reduces the release time and facilitates the scaling process.

Sommario

Questa tesi illustra gli aspetti chiave dello sviluppo e della scalabilità di un sistema di controllo intelligente per integrare più prodotti e siti di produzione, sulla base di un'analisi condotta in collaborazione con il dipartimento di Data Science di Haleon UK. Oggi i consumatori sono alla ricerca di una gamma sempre più diversificata di prodotti per l'igiene orale e in particolare di dentifrici. Questa tendenza verso prodotti sempre più in linea con le loro esigenze determina una curva della domanda in continua crescita. I processi di produzione dei dentifrici sono altamente complessi e comprendono passaggi cruciali come l'aggiunta di ingredienti, miscelazione, il controllo di pressione, temperatura e velocità. Per monitorare in modo proattivo questi processi, Haleon, multinazionale che opera nel settore consumer health care, ha sviluppato un sistema di controllo intelligente quasi in tempo reale che garantisce un prodotto ottimale non solo in termini di specifiche di prodotto ma anche in termini di tempi, costi e risorse dei processi stessi. Nel mondo degli affari di oggi, le soluzioni digitali giocano un ruolo cruciale nel migliorare l'efficienza operativa e nel facilitare l'analisi e la gestione più sofisticata dei dati. Questo progetto di tesi è rappresentativo di un problema industriale standard nel mondo reale, che consiste nel costruire e scalare rapidamente una soluzione digitale accelerando l'adozione e la generazione di valore. Gestire efficacemente una produzione ampia e diversificata richiede che il sistema di controllo sia facilmente e rapidamente scalabile per integrare più prodotti e siti di produzione. La tesi ha lo scopo di identificare gli aspetti più rilevanti affrontati durante il processo di sviluppo e scaling di un sistema di controllo sia da un punto di vista tecnico che gestionale. Da un punto di vista tecnico viene esposto non solo il cuore del modello di data science che è costituito dal Fast Dynamic Time Warping, un algoritmo di machine learning che consente il rivelamento proattivo delle fasi dei diversi processi e siti di produzione, ma anche una panoramica dell'architettura digitale che deve essere robusta e ben progettata per garantire che siano presi in considerazione i diversi bottlenecks dei siti di produzione e i requisiti di tutti i processi coinvolti. Da un punto di vista gestionale, data l'elevata specificità e gli stringenti standards di qualità dei processi di produzione coinvolti, la soluzione

viene sviluppata internamente ponendo come sfida gestionale principale quella di sviluppare e implementare simultaneamente la soluzione con lo scopo di generare valore nel minor tempo possibile. Questa sfida è facilitata tenendo conto dei bisogni di tutti gli stakeholders e dalle accurate e informate decisioni prese durante tutto il ciclo vita del progetto (dal Proof of Concept fino alla fase di scalabilità) ponendo una particolare attenzione sul miminum viable product (MVP), una versione semplificata della soluzione finale contente già tutte le funzionalità chiave che permette di ridurre il tempo di rilascio e facilita il processo di scaling.

Table of Contents

List of Figures

List of Tables

1 Introduction

Haleon manufactures and distributes globally a broad portfolio of glycerol-based oral health and hygiene products. Glycerol-based toothpastes are very complex fluids. In order to guarantee the required quality parameters, the manufacturing process reflects this complexity and is divided into different phases characterized by the addition of various ingredients and mixing in different conditions (speed, temperature, pressure, etc.). Haleon's Data Science Department in collaboration with the Supply Chain Technology department has developed a very innovative nearly real-time control system for the manufacturing process of glycerol-based toothpastes. The control system contains a series of mathematical models and machine learning algorithms implemented in a complex digital architecture, which ensures the execution of the control process in real time.

In the course of this analysis, the goal was to carefully investigate the key considerations that have emerged during the development and scalability of the smart control system for the integration of multiple products and manufacture sites of Haleon (UK). The scalability process represents the final step of a large and ambitious project called "Golden Batch" whose objectives will be explained in a dedicated chapter. To achieve a comprehensive understanding of the reasons that led the company to set the goal of an easily scalable smart control system and the challenges this goal entails, the thesis project aims to explain in the first part the underlying forces, the problem and the proposed solution by tracing the entire life cycle of the project development. In the second part, emphasis will be placed on the technical aspects of the smart control system and my contribution during the model extension to another manufacture site Oak Hill (NY).

1.1 Haleon

Haleon Trading Limited UK is a British world-leading consumer healthcare company with headquarters in Weybridge, England. It has nine large-scale multinational Power Brands and a strong portfolio of Local Growth Brands. Combined, Haleon is positioned to play a vital role for people all around the world, in a sector that is growing and more relevant than ever. The company is a global leader in over the counter medicines with a 7.3 % of market share. Haleon was established on 18 July 2022 as a corporate spin-off from GSK. Sir David Lewis is chairman, with Brian McNamara as Chief Executive Officer. Haleon is listed on the London Stock Exchange and is a component of the FTSE 100, with a secondary listing on the New York Stock Exchange.

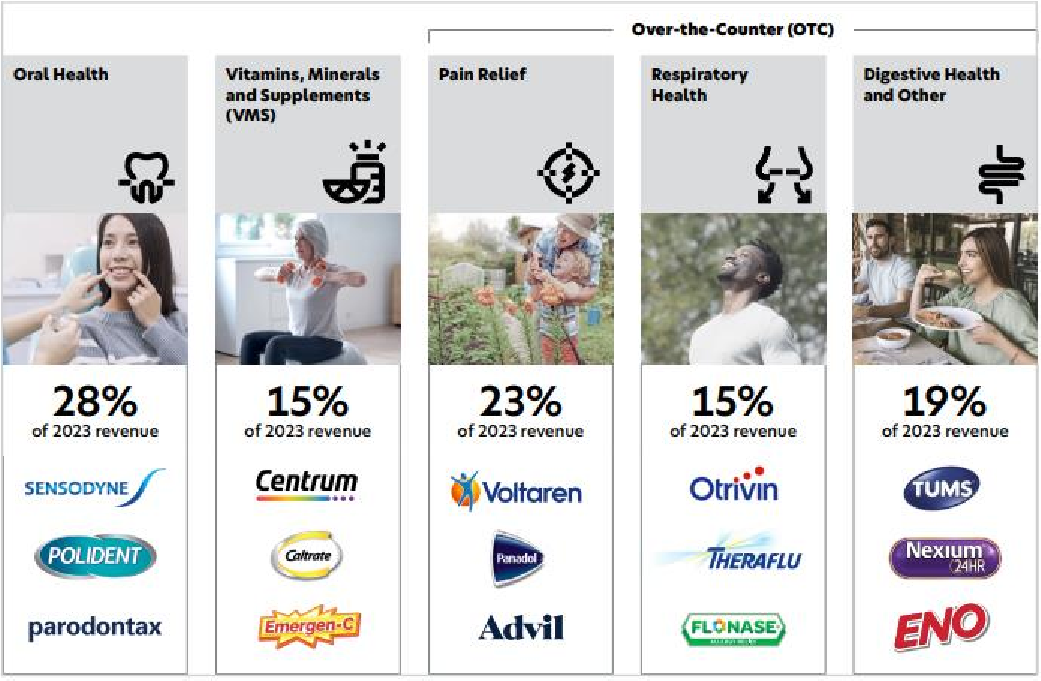


Fig. 1 Leadership position across five major categories by https://www.haleon.com/

Haleon's business strategy is responsible and consists of three closely interconnected areas of interest.

The first is making the everyday health more inclusive which means stimulating research and improving well-being through the solutions they offer.

The second area of interest is reducing the environmental impact. Haleon is committed to operate sustainably by handling the carbon emissions and by exploiting greener packaging practices.

The third area focuses on ethical, responsible and transparent behaviour. It strongly believes that trust with customers, partners and the community is a fundamental aspect. The area promotes social responsibility, the protection of worker's rights, and the promotion of diversity and inclusion within the organization. *"Running a responsible business for Haleon is a strategic imperative"*. [1] [2]

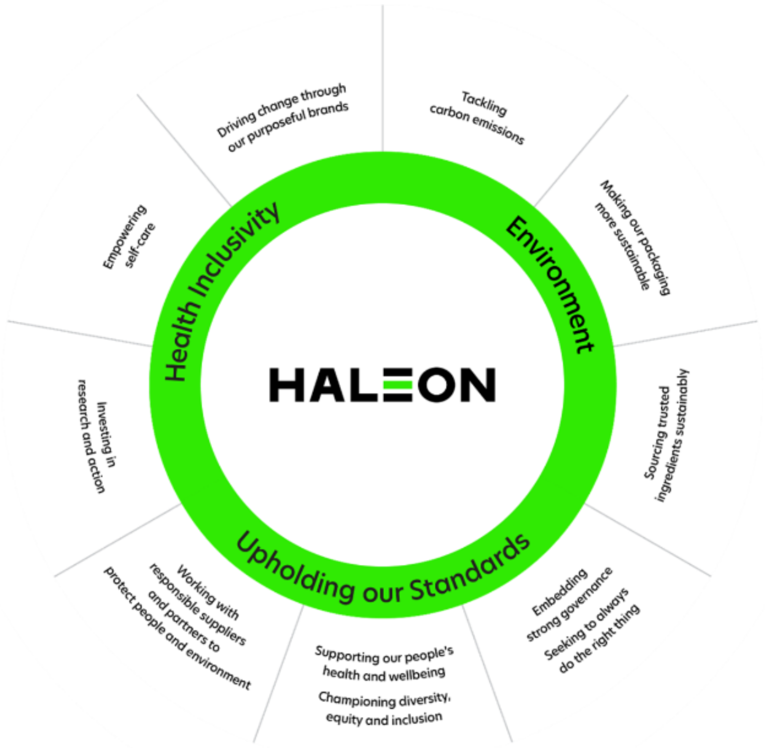

HALEON
Health Inclusivity
Environment
Upholding our Standards
Driving change through our purposeful brands
Tackling carbon emissions
Making our packaging more sustainable
Sourcing trusted ingredients sustainably
Embedding strong governance
Seeking to always do the right thing
Supporting our people's health and wellbeing
Championing diversity, equity and inclusion
Working with responsible suppliers and partners to protect people and environment
Investing in research and action
Empowering self-care

Fig. 2 The three areas of interests of Haleon by https://www.haleon.com/

1.2 Thesis Structure

A small description for each chapter.

- Chapter 1 → It provides a general overview of the context in which the project was developed. The main aspects of Haleon, the company in which the thesis project was developed are highlighted explaining its mission and main achievements.

- Chapter 2 → it focuses on the core problem from which the analysis originates, presenting it as a common challenge for large companies seeking innovation and efficiency in the digital age.

- Chapter 3 → it presents the whole life cycle of the project from its inception, the Proof-of-Concept phase, to the scalability process, highlighting the fundamental role of an efficient Minimum Viable Product (MVP).

- Chapter 4 → This chapter outlines the project management framework adopted for the development of this project, which is based on Agile philosophy. It also explores the key differences between Waterfall and Agile project management methodologies.

- Chapter 5 → it aims to show the digital architecture of the system, focusing on the most relevant aspects for an effective scaling process of the smart control system.

- Chapter 6 → it shows in detail the heart of the Data Science Model which is the Fast Dynamic Time Warping, a machine learning algorithm that allows the near-real time monitoring of the toothpaste manufacturing process.

- Chapter 7 → it explores the key aspects for the extension of the data science model and practical contribution that support the scaling process of the smart control system towards the integration of a toothpaste process from a different manufacture site located at Oak Hill (NY).

- Chapter 8 → it explains the conclusions on the most relevant considerations for strategically managing the development and scaling of a smart control system, based on the personal experience at Haleon’s London office. It provides with practical recommendations and insights that proved to be strategic during the development of this project.

2 Problem Statement

Generally, the industrial processes of the company are characterized by high levels of automation, showing ongoing commitment to efficiency and innovation. However, the current project focuses specifically on areas that show significant opportunities for improvement.

The goal of this paragraph is to declare the problem that characterises these areas that can be optimised leading to an increase of productivity, reduction of costs and an improvement of the output quality.

The areas subjected to analysis are characterized by process steps performed manually by operators who follow specific instructions. The human intervention introduces non-negligible variability, resulting in a variable batch cycle time and, in some cases, the presence of outliers with respect to the product quality specifications. This variability can affect the overall efficiency and quality performance.

The need to address variability and obtaining an optimal performance level leads Haleon's Data Science Department in collaboration with the Supply Chain Technology department to develop a very innovative nearly real-time control system for the manufacturing process of toothpastes. The control system contains a series of mathematical models and machine learning algorithms implemented in a complex digital architecture, which ensures the execution of the control process in near real time allowing an immediate detection of anomalies as result.

In the digital era, *"the only truly sustainable competitive advantage is the speed at which an organization can sense and respond to the needs of its customers. Its strength is its ability to deliver value in the shortest sustainable lead time, to evolve and implement new strategies quickly, and to reorganize to address emerging opportunities better".* [3] Consequently, the data science (DS) model must be scalable as the company operates at multinational level and it needs to handle large quantities of consumers, production volumes and product diversity.

In the consumer healthcare sector in which the company operates especially in the toothpaste production, the processes are characterized by high degree of complexity, specificity, and stringent quality standards. This translates into the need to develop a scalable model in-house, as off-the-shelf solutions from suppliers may not meet the unique requirements. A big management challenge arises: a simultaneous development and implementation of the DS model with the aim of reducing release time and meeting stringent quality standards in line with the company's objective of having an early return on investment.

In the following paragraphs are shown separately the technical and management challenges in order to provide a clear overview of both aspects. The technical aspect includes crucial details that must not be neglected when a smart control system is developed and scaled. The management challenge shows relevant key considerations needed for the implementation of process so that it happens as strategically and efficiently as possible. [3]

2.1 Technical Challenge

The technical challenge to be addressed concerns how to make the smart control system configurable in order to integrate new product and new manufacture sites neglecting no relevant aspects.

In order to ensure that the investment is worthwhile, it is crucial not only to obtain a DS model is valid in performing its functions, but also that it has an infrastructure able to meet essential aspects specific to this type of scaling process. This approach ensures that the system can effectively exploit the opportunities arise in the growth and development of the company.

The most relevant aspects are:

Product and site diversity: The integration of multiple products and multiple sites can lead to a wide range of requirements that must be included in the model. The technical challenge is to achieve a highly flexible system able to accommodate all these requirements without compromise the overall effectiveness. An in-depth analysis on the similarities between different type of products and sites is essential to make easier the scaling process. By exploiting the similarities as a strategic starting point allows to develop a scalable model able to support its extension.

Different types of products

- Aqueous Toothpastes
- Non-aqueous Toothpastes

Haleon Sites

Each manufacture site is characterised by specific features. Different sites can have different bottleneck, multiple types of mixers, and different type of sensors.

- Maidenhead (UK)
- Oak Hill (US)

Data management: to integrate new products and new sites is needed to manage a large amount of data coming from different sources. An example of data could be the process specification measurements such as temperature, pressure, different type of speeds… that are constantly collected per each toothpaste coming from different sites. The challenge is to have a robust infrastructure able to support this high volume of data ensuring that is available in a usable way to support the decision-making process.

The complexity of the user interface: The expansion of the means including more and more information about a higher number of products and manufacture sites. This growth leads to more complexity in the user interface because is crucial to include all relevant aspects related to each new guest in the system. The user interface has a diversify audience that includes workers with different set of experiences and skills such as operators, technicians, product owners, data scientists, managers… This is an aspect that should not be overlooked because the level of complexity of the user interface must be such that it allows everyone to have a clear and defined understanding of the performance of the process so that all aspects are taken into account.

2.2 Management Challenge

Given the high specificity of the inherent toothpaste production process, one of the most management challenge of this project lies in the fact that it needs to be done in-house. It implies that it being simultaneously developed and implemented in order to generate value promptly and obtain an early return on investment.

To efficiently handle this journey toward smart control system scalability, it is essential to match the needs of all key stakeholders who hold roles in the most relevant and crucial areas of the project. Each stakeholder has specific expectations that have to be considered to create a scalable solution that are effective and sustainable. The most relevant stakeholders or areas of greatest interest are operators, digital architecture, process engineering, and data science.

- **Operator involvement:** a strategic starting point could be involving operators to collect feedback as they have a direct insight on the operational needs. Their practical experience is valuable for understanding how the changes will affect the day-to-day work. Operators can suggest things that might not be promptly clear at management level allowing to correctly identify potential improvements and ensuring that solution is sustainable in the everyday operational context.

- **Digital Architecture:** it provides the technological foundation that if well-structured allows the system to be scaled easily and rapidly, ensuring that all part work together in an effective manner. The importance of the Digital Architecture leads me to explain in a dedicated chapter this area in a more detailed manner.

- **Process engineering:** the process engineers help to assess that the performance monitoring of new processes into the model is valid and reliable based on the information they hold on process performance. The aim of their work is to ensure the toothpaste quality by means of rigorous quality controls during all steps of the manufacturing process.

- **Data science**: it represents the heart of this project because it is the pillar on which analysis and interpretation of data is based. Thanks to advanced analysis, data science allows to obtain meaningful insights and make informed and reliable decisions. When the aim is to scale a smart control system, one of the most critical aspects is the adaptation of the data science model to ensure that it can fully meet the requirements of new products and sites. A well-designed data science model must be flexible and robust able to accommodate easily new guests without compromising the overall effectiveness.

3 Proposed Solution: ‘’Golden Batch Project”

3.1 The Goal of the Project

The proposed solution to the problem explained above is an extensive project called “Golden Batch”. *What is meant by Golden Batch?* It is a special batch that meets all the highly stringent quality standards that these specific toothpaste manufacturing processes require. The basic requirement is to have an accurate anomaly detection and pattern recognition techniques able to recognize process deviations at early stages. Thanks to the use of machine learning algorithms and advanced mathematical modelling a “golden batch” manufacturing patterns will be detected able to guarantee an optimal configuration from a quality and batch cycle time requirements.

This approach guarantees the quality of the output product leading to a significant reduction of the scraps and rework, while optimizing the overall operational efficiency. Over the course of these months, some team members dedicated on this specific work relating to quality, covering this crucial role in ensuring that every manufacturing step complain with required quality standards with a particular attention to more priority features.

The Golden Batch target is not only for a particular process or manufacture sites, but it is the final goal that includes all the products and sites under the control of the Data Science team of Haleon. This is the reason why one of the first goals is to achieve a Data Science Model that is easily and quickly scalable to multiple products and sites.

3.2 From PoC to MVP: toward Scalability

The past decade has seen a significant increase in interest and adoption of innovative methodologies that emphasize agility and the ability to react quickly to the market needs.

The first aim of these methods is to optimize investments by identifying in a shorter time frame which strategies are effective and which should be discarded trying to constantly keep up with innovation.

Going through the project development life cycle is a useful experience that provides a clear and comprehensive view of where we are today, what was the starting point, and what is expected to achieve in the future. The path from the launch of an idea to the implementation of a solution consists of crucial decisions that influence and determine the success of the project and its sustainability. During my thesis project development, I had the opportunity to understand what the most important life cycle steps from a business context standpoint are.

The essential phases of the development life cycle of this project are:

- The Proof of Concept (PoC);
- The Prototype;
- The Pilot phase;
- The Minimum Viable Product (MVP);

At the end of these stages, it is possible to find the **scalability**. Every phase has a distinct purpose that serves as a guide in the development of this project.

This paragraph aims to clarify these concepts by adapting them to the specific project developed by Haleon. [4]

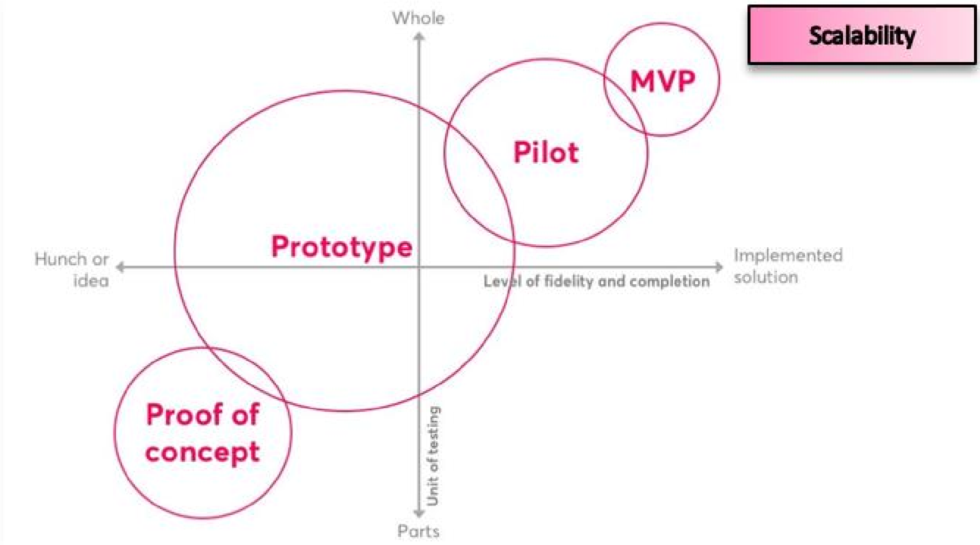


Fig. 3 Different phases toward scalability by https://www.nesta.org.uk/blog/proof-of-concept-prototype-pilot-mvp-whats-in-a-name/

3.2.1 Proof of Concept (PoC)

The Golden Batch project initially needed to go through a Proof-of-Concept phase due to its innovative and thus uncertain nature. The POC is a preliminary stage where ideas begin to take shape. It involves research and exploration where theoretical approaches are tested to validate if they can work in a practical environment.

The point of the Proof of Concept does not involve analysing every detail of the idea but focusing more on the aspects that are most likely to generate the highest probability of criticisms.

For example, in this case the focus firstly was on the feasibility of the predictive model and its ability to perform what is expected by investing time and resources on the more uncertain points that characterized the project.

It can be deducted that the objective of this `beginner phase is to mitigate potential future risks and provides an early assessment that serves as a guide for the future decision-making process. In this specific case, moving to the next phase was enabled by having demonstrated that developing a golden batch and implementing a scalable model was feasible in its business context. [5]

3.2.2 Prototype

After passing the feasibility on an idea through the step shown above, the next step involves the development of a prototype.

At this point, the focus shifts from feasibility to usability and effectiveness. Unlike the initial exploration phase, this is more focused on features that enable implementation in the real-world context. This phase is critical because it will allow to move towards the next stages that will require significant investments of greater relevance.

The success of prototype is measured by its ability to meet functionality objectives of the proposed solution. This phase requires the involvement of additional parties enabling the development and future implementation differently from the Proof-of-Concept stage. Because of this broader involvement of actors, prototyping requires more time and resources than the previous stage, but it is still less expensive than the full-scale system development.

In the Golden Batch project, the prototype was a predictive model built firstly on a specific glycerol-based toothpaste manufacture process for a specific production site (Maidenhead, UK). The reason why the prototype was exactly constructed on this specific manufacturing process and site will be explained in the course of this analysis. [5]

3.2.3 Pilot

The next step is to launch a pilot project, after the prototype has been improved and considered as effective. The current step is where the solution is tested in a real-world scenario, even if on a smaller or more controlled scale than the full-scale implementation intended to be in the future. It is critically important for building trust among stakeholders and end users.

This phase lays the foundation for scalability while ensuring that the model can actually be effective in the business context. The main goal is to identify any issues that are more difficult to predict and to gather as much information as possible on its effectiveness. This is a critical phase to get a clearer and comprehensive understanding of the model and ensuring that it is align with the overall objectives of the company.

In the Golden Batch project, the pilot phase is considered to be the more extensive version of the model in which the most essential requirements for accommodating new products or manufacture sites have been included. The success of the pilot phase is measured by the ability to expand the solution without substantial issues. It is still less expensive and costly than full-scale implementation. [5]

3.2.4 Minimum Viable Product (MVP)

Following the life cycle of the project, after the pilot project there is the Minimum Viable Product (MVP). The MVP is a simplified version of the final model that has only the essential features, but it is already able to satisfy the most relevant needs. Minimum Viable Product is a concept used in the business context that is adopted as a strategic approach to optimize the development time and costs. This approach allows the company to improve and iterate the model incrementally, basing changes on feedback and actual data collected, by avoiding to develop something that will not meet the expectations.

In the Golden Batch project, The Minimum Viable Product is the configurable version of the predicted model.

For the development of my thesis project this phase holds a different relevance to the others because it was my starting point. I joined the Data Science team at a time when a Minimum Viable Product had just been developed and the most important focuses were how to improve the configurable version and scale the model so that multiple products and sites could be properly monitored. This phase allowed me to deeply understand the importance of the minimum viable product and to actively contribute to the optimization and expansion of the system.

In the following paragraph 3.3 it will be explained the reasons why it was decided to obtain a particular minimum viable product, on which the scaling process then started. The detailed and strategic reasons behind this decision will be explained by highlighted how the MVP facilitate the achievement of the key objectives and support the expansion of the model for additional products and sites. [5]

3.2.5 Scalability

Scalability represents the final phase of the entire project life cycle that follow the minimum viable product phase.

Once a concept has demonstrated its feasibility in the Proof of Concept, its usability thanks to the prototype, its launch in the pilot phase, and after having achieved the essential features in the minimum viable product, it is ready to be refined and scaled.

This phase aims to ensure that the model can meet the high amount of data, production processes and product diversity that characterize its multinational nature. Achieving an easily and rapidly scalable model means, in this specific context, monitoring and controlling multiple products and other productions site allowing the company to have greater operational efficiency, more reliability of product batches and to be aligned with the overall objectives.

3.3 The MVP choice

This paragraph explains the reasons why a particular minimum viable product was chosen. Generally, the choice of a proper MVP is crucial for quickly testing and adaptation of the model to the market needs starting from an initial version that already contains all the key functionalities.

In this specific analysis which focuses on the identification of the most important aspects to manage the scaling process in the most efficient manner, the choice of the minimum viable product is crucial especially because of the management challenge this project requires: to simultaneously develop and implement the solution with the aim of reducing release time and generating promptly value in order to have an early return on investment.

A careful choice of MVP enables the above mentioned management challenge to be addressed efficiently allowing for an effective base that is well-functioning and at the same time easily adaptable to other processes and production sites so that it can generate value promptly making the time to market faster.

In other words, the choice of an accurate minimum viable product is a strategic move to generate immediately value with the current version and quickly adapt it according to the needs of the near future.

The MVP was a configurable version of the model built on specific glycerol-based toothpaste manufacturing processes from the Maidenhead (UK) site especially for the following reasons:

1) *High production*

Maximum impact → based on the processes that hold the highest percentage of production ensures an immediate generation of value since are optimizing processes that will significantly affect a largest slice of the entire production system.

2) *The longer Batch Cycle Time of the Glycerol-based Manufacturing Processes than Aqueous toothpaste.*

- Maximum impact → to focus on the manufacturing toothpaste processes with the highest batch cycle time enables to address critical aspects at the early stages obtaining substantial positive impacts on the overall production timeline.

- Learning opportunities → working firstly on the products with the highest batch cycle time provides the opportunity to gain in-depth knowledge of the most challenging manufacturing processes. This knowledge can be useful when the solution needs to be extended to similar toothpaste processes.
- Risk Mitigation → working firstly on the most challenging processes helps to reduce unforeseen issues when extending the model.

The choice of starting from the highest batch cycle time processes set the foundations for an effective and easily scalable model.

3) *The most expensive toothpaste.*

Maximizing the economic impact → focusing the development of the MVP on the glycerol-based toothpastes allows to reduce their defective batches leading to a reduction of their according costs. Being at the same time one of the most expensive products, it inevitably allowed money savings. This choice leads to a substantial improvement of the overall profitability.

4) *The manufacturing processes of all non-aqueous products are similar.*

Easily adaptable → many other products that are intended to be integrated into the system show many common features with the selected non-aqueous products making the model quickly adaptable to products to integrate in the near future.

4 Agile Vs. Waterfall: Which Project Management Methodology is best for the “Golden Batch Project”?

4.1 Project Management Framework

This chapter aims to outline which is the most appropriate project management framework for this project. The broader range of available frameworks provides to the project managers the flexibility to choose the approach that best meets the requirements of the project. In the initial part, it is provided an overview of what is the project management, and which are its the essential phases. As was mentioned in the Chapter 2, the most relevant management challenge lies in effectively managing the development and scaling of the control system which must occur simultaneously in order to accelerate the adoption and delivery value.

A project management methodology is a set of principles required to successfully manage a project. Each methodology is characterized by its own processes and procedures based on its principle.

Each project is characterized by five essential management phases regardless of the management principle adopted, each of them plays a crucial role in the achievement of the objectives within the established time and cost constraints. The above-mentioned stages are:

- Project Initiation
- Project Planning
- Project Execution
- Project Monitoring & Controlling

- Project Closure

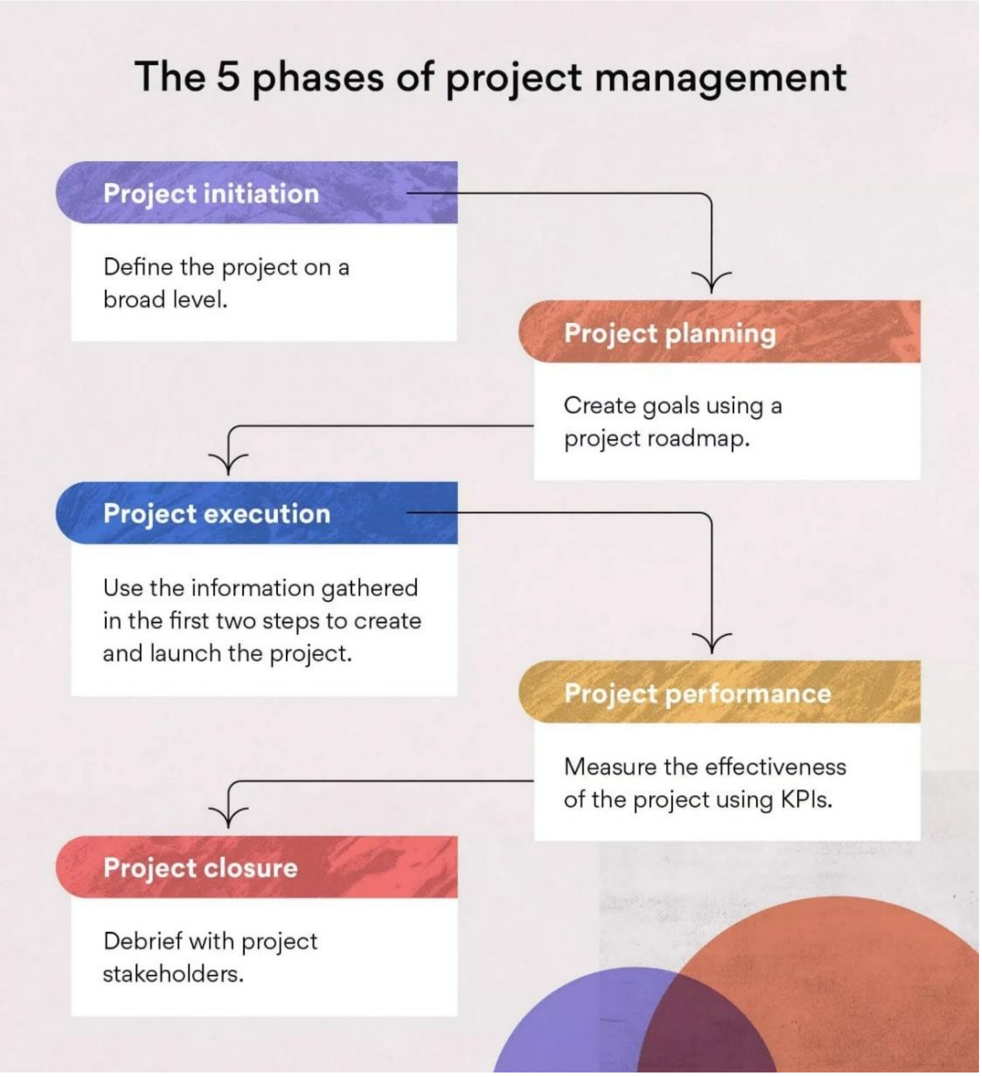

The 5 phases of project management
Project initiation
Define the project on a broad level.
Project planning
Create goals using a project roadmap.
Project execution
Use the information gathered in the first two steps to create and launch the project.
Project performance
Measure the effectiveness of the project using KPIs.
Project closure
Debrief with project stakeholders.

Fig. 4 The 5 phases of Project Management by https://asana.com/resources/project-management-phases

There are two main philosophies that guide the project management methodologies: Waterfall and Agile.

Waterfall Project Management is a well-established project management methodology. It is characterized by its linear and sequential structure. The methodology consists of several phases each of which is built on the final results of the previous phase. Therefore, it is necessary to fully complete one phase before moving on to the next.

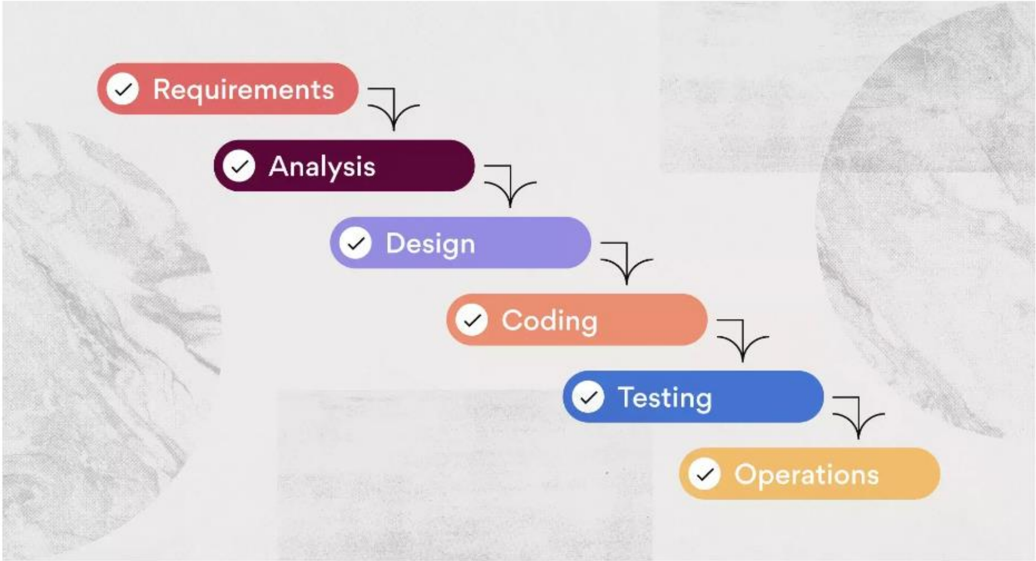


Fig. 5 Waterfall Project Management by https://asana.com/resources/project-management-methodologies

Agile Project Management was design especially for the software development project management that requires greater flexibility and communication. Agile more than being a methodology is a principle. It is based on principles defined in the Agile Manifesto written in 2001 by a group of software developers. This document emphasized the importance of:

- Individuals and interactions over processes and tools.
- Working software over comprehensive documentation.
- Customer collaboration over contract negotiation.
- Responding to change over following a plan.

[6]

The Agile Project Management principle is preferable for this project due to its ability to manage uncertainty and to adapt quickly to changes while ensuring that the final version of the model is more aligned with the objectives of the project. The dynamic nature of this project leaves no doubts as to which basic management principle to rely on.

"With origins in software development industry, it has increasingly been adopted by organizations in a variety of industries experiencing dynamic and risky project environments where the traditional waterfall project management framework fails to prove its effectiveness." [7] Basically, the process is based on iteration cycles with a typical duration of 2 weeks starting from planning phase where all activities are planned, with the aim of adding incremental value at the end of each iteration. Every iteration release at the end a deliverable that can be a new usable functionality of the model. This approach allows for a rapid and continuous delivery that perfectly meets the Golden Batch requirements.

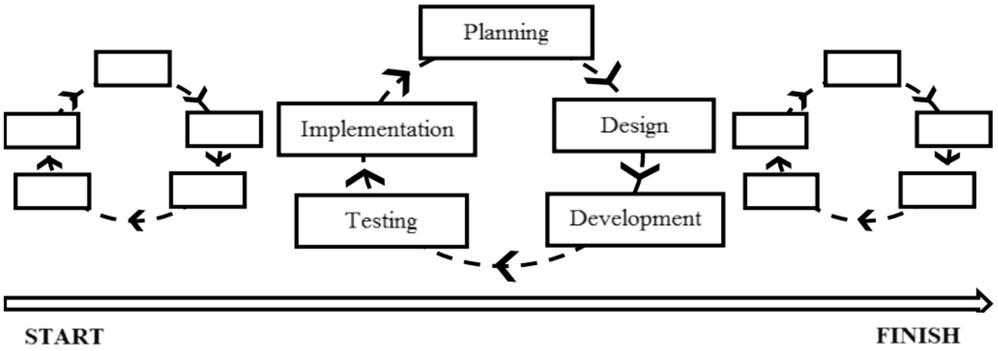


Fig. 6 Agile Project Management by [7]

4.2 The choice

The real challenge lies in the choice of the specific approach to be used in combination with the Agile principle. There are various methodologies such as Scrum, Kanban, crystal, Scrumban and others, each with its own specific characteristics. The choice of the most suitable methodology requires careful assessment of the specific needs of the project, the level of uncertainty, complexity, and others. The combination of the Agile philosophy with a specific approach allows to achieve optimum results. [8]

The methodology chosen for this project combines Scrum and Kanban to leverage the strengths points of both, resulting in a hybrid approach called Scrumban. The two methodologies are firstly explained separately, and then the reasons that guided toward this final choice are illustrated.

- **SCRUM**

  Scrum is one of the most common approaches applied in conjunction with the Agile framework. It is designed to improve collaboration and flexibility especially in the software development.

  The work is divided in *sprints* which are fixed cycles with a typical duration between 1 and 4 weeks. Each sprint starts from a *Sprint Planning,* during which the work for this period is defined. In sprint planning, the *Product Owner* together with the team select from the *Product Backlog* (a list of all activities needed for the success of the project) the activities to be performed for the incoming sprint. During the sprint, the team works on the selected activities and participate in daily meetings called *Daily Stand-up* to synchronize work and fix problems. At the end of the sprint, the team show the work to stakeholders during the *Sprint Review* and identifies potential improvements in the *Sprint Retrospective*.

  The scrum approach defines roles such as Product Owner who has to manage the Product Backlog and ensure that the priorities are respected. The sprints have fixed durations. The essential feature of scrum is the iterative and incremental development of the project with a focus on continuous feedback during the review phase allowing all requirements are met and constantly improved.
  The scrum approach fits well with this type of project that need to take small steps at a time while generating value continuously but that is directed

toward larger goals. In addition to that, it allows to maintain constant collaboration with all key stakeholders involved in the project. [9]

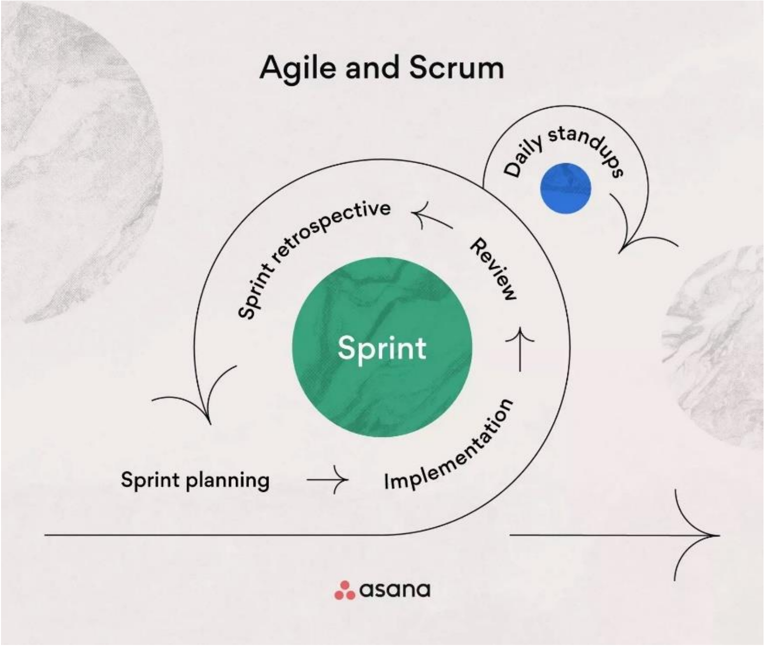


Fig. 7 Scrum approach by https://asana.com/resources/waterfall-agile-kanban-scrum

- **KANBAN**

Kanban is a methodology that helps improve the workflow visibility. Originally developed by Toyota in the 1940s as part of the just-in-time (JIT) production system, and then it has become a common practice not only in manufacturing but also in the software development. The Kanban principle is based on early delivery, adaptive planning and continuous improvements.

This approach uses *Kanban boards* to display all the activities. The board is divided in columns representing the different stages of the work process that include the individual task for each stage. As the work progresses, each task represented by a visual card moves from one column to another according to its stage. Generally, the stages are:

- To do→ Activities that are to be started.
- In progress→ Activities currently in progress.

- Review→ Completed tasks that currently are subjected to review or testing.
- Done→ Completed tasks.

This methodology is also useful to limit the overload of activity by imposing a limit on the number of activities that may be present in the work in progress column making the workflow consistent and improving the delivery value. It encourages daily meetings as the stand-up meetings to assess the progress and identify potential problems supporting a continuous improvement. Kanban approach is characterized by a higher level of flexibility with the possibility to add, remove or modify activities at any time in order to adapt them to changing priorities. [10]

- **SCRUMBAN**

The Scrumban represents the choice of management framework for the Golden Batch Project. It combines elements of Scrum and Kanban taking the best from both to create a hybrid approach. This method is particularly useful when the aim is to keep the structure and discipline of Scrum in addition to the flexibility and continuous visualization of work offered by Kanban.

The approach adopted for this project uses *sprint* cycles to plan the work with a fixed duration of 2 weeks maintaining the scrum meetings such as sprint planning, daily stand-ups, sprint review and sprint retrospective. This decision is taken by the need to maintain a high level of collaboration between all key stakeholders involved in the project without ever losing sight of views and needs of each of them.

In addition to the singular scrum method above mentioned, this hybrid approach allows to add new activities to the work plan during the sprint. This flexibility coming from the Kanban principle enables rapid adaptation to new priorities or emergencies without having the need to wait the end of the sprint. As the project alternates between phases characterized by high unpredictability and others where the requirements are more defined, the best solution is to adopt Scrumban. In other words, Scrumban is a perfect combination between the structure and discipline of Scrum with the flexibility of Kanban allowing the team to quickly adapt to changes during the initial, less predictable phases (such as Proof of Concept) and also handle more defined stages (such as Scalability) in later phases.

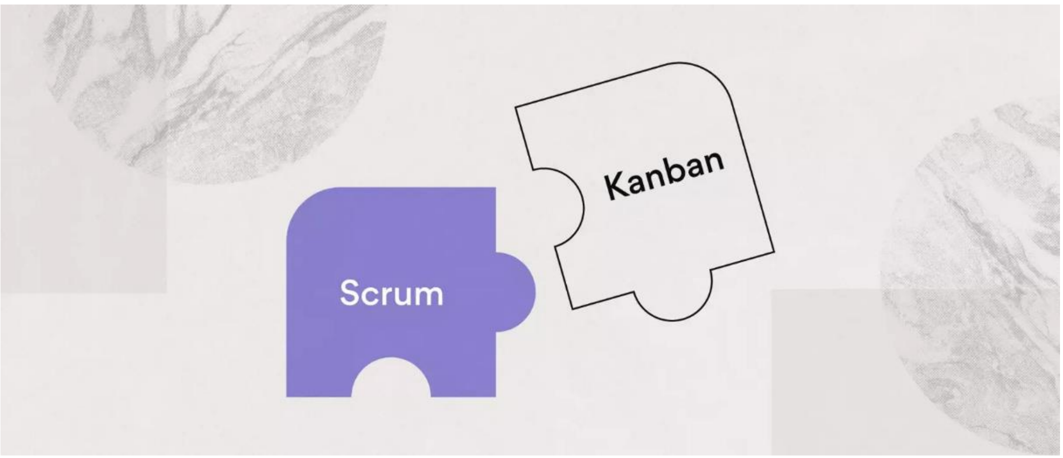


Fig. 8 Scrumban approach by https://asana.com/resources/project-management-methodologies

5 Digital Architecture

The digital architecture is the most important aspect for this type of application, and it is a crucial aspect that allows the company to remain competitive in today’s industrial context.

The digital architecture has to be designed in order to meet some important requirements, such as:

- ✓ Low latency
- ✓ High reliability
- ✓ Scalability

The demand for diversified toothpaste is increasingly high so that the architecture needs to be able to support this growth rate. It is essential that it allows the easily integration of different type of products and manufacture sites safeguarding scalability and flexibility.

The growing demand for diversified toothpaste must always be accompanied by compliance with all the stringent quality requirements. This scenario requires the ability to produce large amount of products while keeping high standards of quality. This type of challenge can be faced through a system that allows operators to monitor in real time the performance of each batch allowing them to detect immediately potential anomalies.

The key for successfully achieving it is to design a robust architecture able to easily accommodate new products and additional manufacture sites. Only with an excellent system the company can strengthen its competitive position while ensuring a successful and sustainable future.

5.1 Interconnected components

In this paragraph the various component of the digital architecture and how they interact each other are explored. Here, it is showed an analysis on their roles and how they work together to create an efficient system.

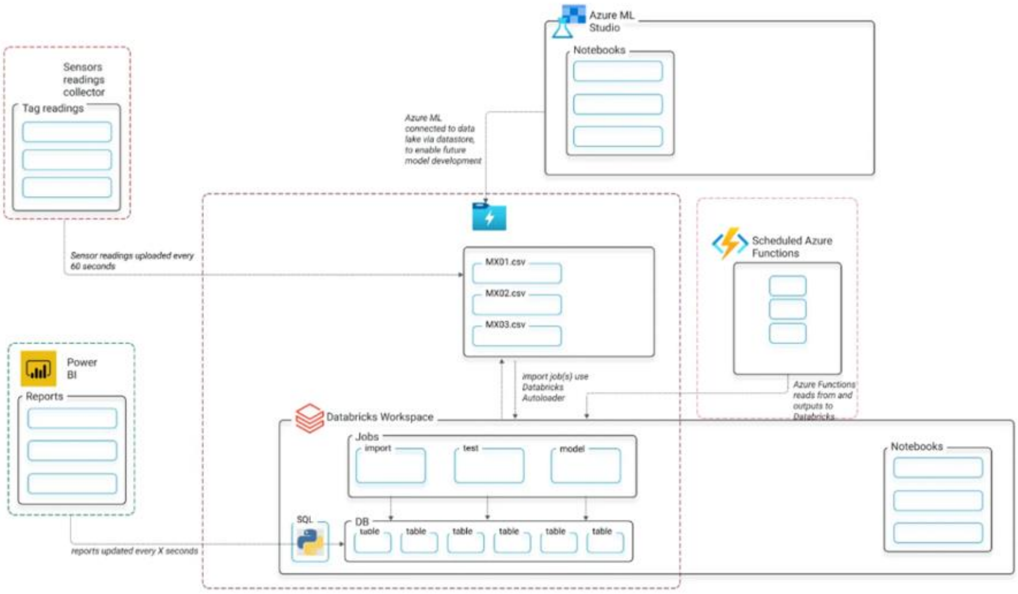


Fig. 9 Digital Architecture by https://medium.com/trusted-data-science-haleon/unleashing-the-power-of-real-time-machine-learning-to-accelerate-the-production-of-toothpaste-in-70ab7c3cf1f1

1. **Sensor readings collector**: Inside the mixers, there are sensors that collect information on critical parameters such as temperature, pressure, weight, etc., which significantly affect the quality of the final product. This system collects readings at a specified frequency.

   It is important to highlight that each sensor is designed for a specific application because each of them requires different accuracy, cost, and environmental impact requirement. Each sensor is designed to meet specific needs. For example, high-accuracy sensors could be applied in medical applications where is highly uncommon to use cheaper sensors. The choice of sensors is particularly important for any application because it directly impacts on measurements. A suitable choice guarantees that the system performs as intended with the respect of the overall efficiency.

   In addition, each sensors performs accordingly with the context in which it is placed. Several factors affect the sensor performance, a common example might be the environmental factors. As will be demonstrated in the next chapter 7, during the explanation of the task 2 performed by me, different mixer contexts lead to different sensor readings. For instance, the same system located in different mixers can produce distinct data measurements required additional in-depth analysis.

2. **Azure Data Lake**: it is a centralized repository that stores large volumes of data in its original form. In this specific context, it receives measured data from sensors at regular intervals, these are stored in a format that allows processing in later stages to be facilitated. [11]

3. **Databricks Delta Live Tables**: This component extracts, transforms and loads the raw data from the data lake and put them into a structured format saving them in the resulting tables. This process is carried out real-time. This dynamic nature ensures that the data is always up to date and readily available for predictive analysis.

4. **Azure Functions**: databricks tables are kept updated with the corresponding predictions by means of the azure functions: these functions run the predictive models and loads those predictions into the databricks tables.

5. **Power BI**: real-time data are visualized by the Power BI. This tool allows to gain insights from the visualized data to detect anomalies and apply corrective actions to preserve the optimal functioning of the system. [12]

5.2 Key Aspects of the Scaling Process

One of the essential elements of the Digital Architecture is the Azure Functions, which is used to execute efficiently this model with the ability to do it in parallel. In other words, it provides with the ability to run several operations simultaneously. Azure Functions are set up to make forecasts at regular intervals, allowing the system to update them frequently. The parallelism enabled by these functions makes it possible to work on several mixers concurrently, which is essential for this type of scaling process. Of course, all the interconnection of elements involved is crucial in order to achieve a digital architecture that maximises operational efficiency and adapts to changing production conditions.

The costs associated with a digital architecture capable to support such work are significant and it can begin to weight on company budgets. The more advanced the technological infrastructure is required, the greater is the investment. When the aim is to integrate complex components that support this scaling process, it is important to assess costs and benefits arises, in order to ensure that the investment is worthwhile and guarantees a long term sustainability.

6 Data Science Model

Toothpaste manufacturing processes are very complex with the ultimate goal of obtaining a product that ensures that all rigorous quality standards are met. The most critical steps that ensure the optimal formulation of toothpaste are the addition of ingredients to the mixer, temperature adjustment, pressure change, speed adjustment (speed of the scraper, agitator, and homogenizer), and cooling activity.

The high specificity of the process is threatened by the manual nature of production, characterized by inefficiencies and delays during the execution of each step. More often than not, these delays are due to the lack of real-time visibility and monitoring, with the risk of impacting the overall efficiency of production processes.

Operators must manually monitor the progress of various features to ensure the completion of each phase. This manual monitoring not only introduces the potential for human error but also limits the precision of the process. Therefore, the Golden Batch project aims to address these challenges by developing a smart control system easily configurable to all toothpaste variants and manufacture sites of Haleon. One of the main goals of the Golden Batch project is the live identification of steps throughout the toothpaste manufacturing process. By providing a comprehensive solution to the challenges associated with manual production processes, the Golden Batch project aims to improve the reliability, consistency, and quality of toothpaste production.

Haleon's Data Science Department has developed a smart control system for Non-Aqueous toothpastes manufacturing processes. This control system contains mathematical models and machine learning algorithms integrated within a sophisticated digital architecture, guaranteeing real-time monitoring of the glycerol-based toothpaste manufacture process. The aim of the Data Science Model is to recognize all process phases from machine sensor readings in real time.

The identification of the steps results to be extremely complex due to the intricate nature of such processes. The variety of ingredients used, along with frequent changes in temperature, pressure, and speed, make it difficult to detect each phase of the process. This complexity is further heightened by the need to maintain high standards of quality and safety during production. Therefore, accurately identifying the steps in the process represents a significant challenge. **Fast Dynamic Time Warping** (FastDTW) is used as a detection method.

This chapter offers an in-depth exploration of Fast Dynamic Time Warping and its application within the manufacturing context of Haleon. Initially, it delves into the intricacies of FastDTW, elucidating its principles, algorithms, and computational techniques. Through detailed analysis, it provides a comprehensive understanding of how Fast Dynamic Time Warping serves as a valuable tool for monitoring and optimizing manufacturing processes.

Furthermore, the chapter explores key aspects related to the Data Science Model extension to monitor additional processes and manufacture sites. Through this exploration, it becomes evident how scalability is achieved. Lastly, it concludes by highlighting my contribution related to the extension phase of the model.

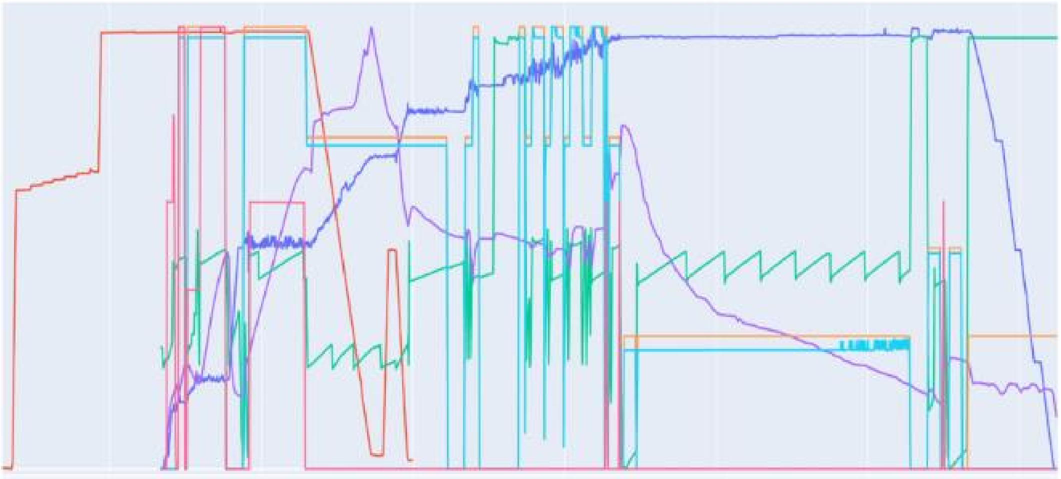


Fig. 10 Example of complexity of toothpaste manufacturing process

6.1 Fast Dynamic Time Warping

Dynamic Time Warping (DTW) is a technique used to compare two temporal sequences of varying lengths and find the optimal alignment between them by warping one sequence in time relative to the other. It is often used in fields like speech recognition, gesture recognition, and time series analysis. The goal is to find the optimal alignment between the two sequences, even if the timings between them are different.

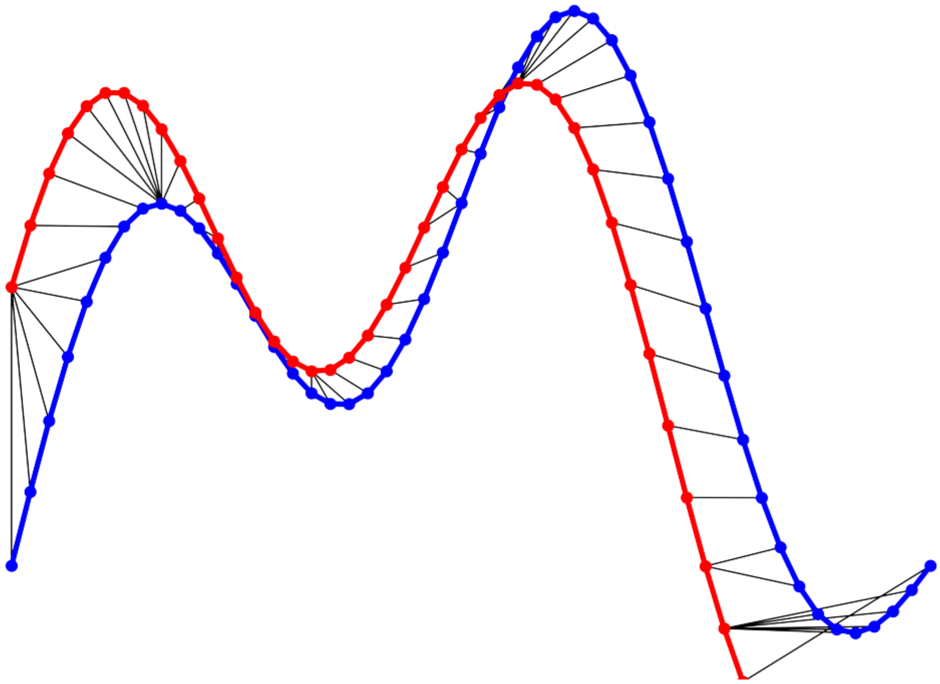


Fig. 11 Dynamic Time Warping by https://ealizadeh.com/blog/introduction-to-dynamic-time-warping/

FastDTW is an approximation algorithm to compute DTW more efficiently and in particular is suitable for large data sets or sequences. Traditional DTW has a time complexity that makes it computationally expensive for long sequences. FastDTW aims to reduce this complexity while still providing reasonably accurate alignment.

"This efficiency is achieved through down-sampling, which reduces the number of points involved in distance calculation, and a constraint on the warping path, limiting the number of cells on each side of the path." [13] It decomposes the

sequences into smaller segments, calculates the DTW on these segments, and then combines them to find the overall alignment. This hierarchical approach significantly reduces the number of comparisons needed, thus speeding up the process. However, in scenarios involving short time series, DTW can outperform FastDTW in terms of speed.

Although FastDTW sacrifices some accuracy compared to the exact DTW algorithm, it often provides satisfactory results in practice while being much faster, making it suitable for real-time or large-scale applications where computational efficiency is critical.

In this project, FastDTW enabled the efficient alignment of procedure models for toothpaste production in real-time processes and the accurate detection of production steps.

6.1.1 How the DTW algorithm works.

Dynamic time warping (DTW) works by calculating the similarity between the 2 sequences, allowing you to measure how well one sequence fits another. This is done by finding the best match between points in the 2 sequences, allowing for a temporal deformation of sequence to maximise similarity.

Before explaining the details of how the algorithm works, it is important to highlight that there are several approaches to calculating the distance that are distinct and useful for comparing and aligning data sequences. The Euclidean distance serves as a direct measure of the distance between two points in Euclidean space. This approach directly compares point-by-point sequence values without considering time alignment differences. It is sensitive to the scale of data and does not take into account variations between sequences. On the other hand, DTW is an algorithm that aligns two time sequences considering the differences. It offers greater flexibility than the Euclidean distance in that it allows non-linear alignments between sequences, adapting to variations.

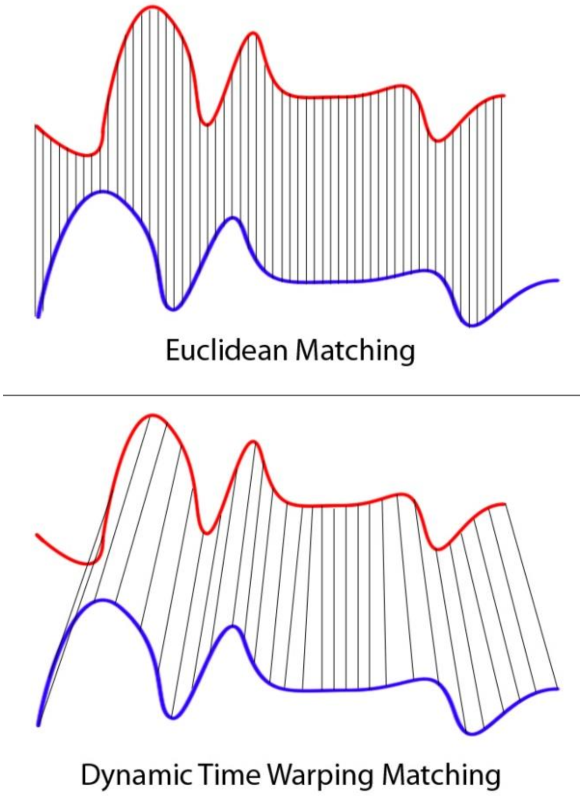


Fig. 12 Euclidian vs DTW by
https://commons.wikimedia.org/wiki/File:Euclidean_vs_DTW.jpg

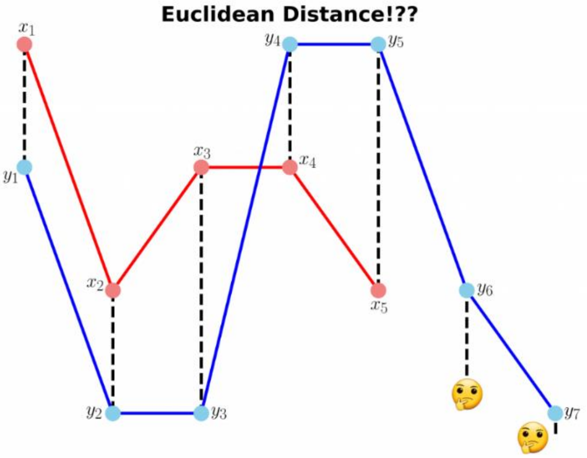


Fig. 13 Euclidean Distance by
https://ealizadeh.com/blog/introduction-to-dynamic-time-warping/

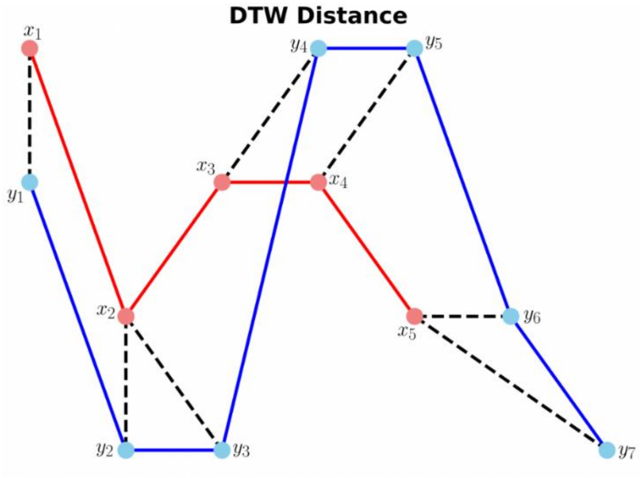


Fig. 14 DTW Distance by https://ealizadeh.com/blog/introduction-to-dynamic-time-warping/

In summary, while Euclidean distance directly compares sequence values without considering timing, DTW matching takes into account variations, finding an optimal alignment between sequences. DTW is better suited for sequences with distortions. In a real context, the Euclidian distance is often used for determining time series similarity; however, where series are out of phase, DTW can be much better method.

Another clarification to make is the difference between distance and similarity. Similarity is a concept that tells us how similar two sequences are. Meanwhile, the distance is the tool we use to evaluate the similarity: the smaller the distance, the greater the similarity between the two sequences. [14]

In the following, is provided an overview of the algorithm and notation. There are 2 sequences: sequence X and sequence Y. Sequence X consists of N points, while sequence Y consists of M points. The goal is to align the 2 sequences. This will be accomplished using a cost matrix (D).

- Inputs: $x_{1:N}$ and $y_{1:M}$
- Cost matrix: D ∈ $R^{N+1 \times M+1}$

```
Initialization:
    for i = 1 to N: $D_{i,0} = \infty$
    for j = 1 to M: $D_{0,j} = \infty$

Calculate cost matrix:
    for i = 0 to N:
        for j = 1 to M:
            $D_{i,j} = \mathrm{d}\,(x_i\ , y_i) + \min\,\{(D_{i-1\,,j-1}), (D_{i-1\,,j}), (D_{i\,,j-1})\}$
```

Get alignment: Trace back from $D_{N,M}$ to $D_{0,0}$

This cost matrix helps to identify the corresponding points between sequence X and sequence Y, as well as the cost associated with matching these points. The matrix size is (N+1)x(M+1). In other words, it will have as many rows as there are points in the X sequence plus 1, and it will have as many columns as there are points in the Y sequence plus one.

- The inputs are the 2 sequences being analysed.
- The cost matrix is crucial for identifying the minimum cost path for aligning the sequences in the most cost-effective manner.

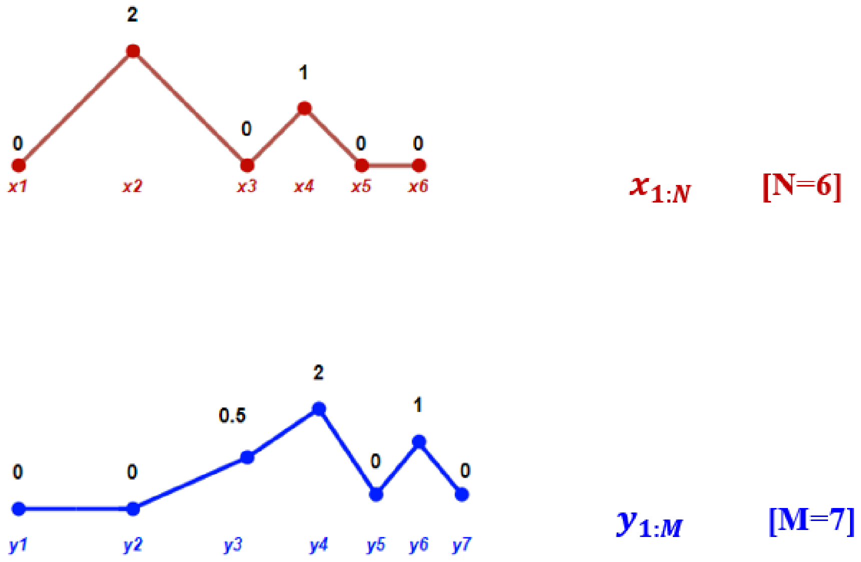


Fig. 15 Inputs

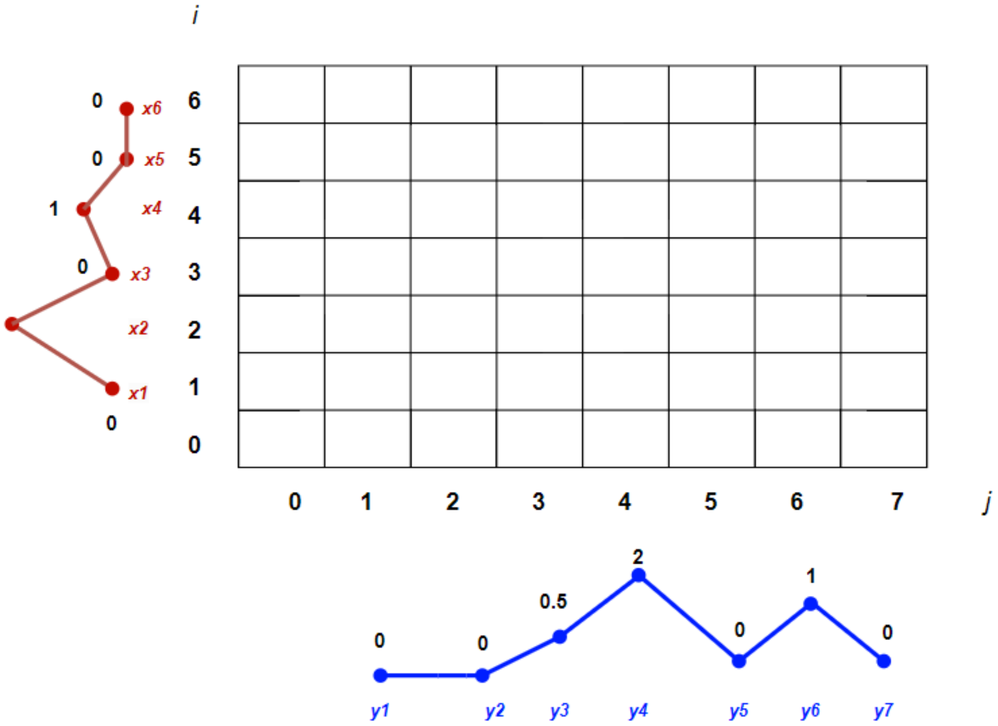


Fig. 16 Cost Matrix

As shown in the above image, the rows (i) of the matrix are numbered in an unconventional way (from the bottom to the top), whereas the columns (j) are numbered in the usual manner. Once the inputs are defined and the matrix is constructed with the correct dimensions, the next step is to initialize the matrix.

- Initialization:

  for i = 1 to N: $D_{i,0} = \infty$

  for j = 1 to M: $D_{0,j} = \infty$

In other words, the initialization requires filling the entire first column and the entire first row with ∞ except for D(0,0).

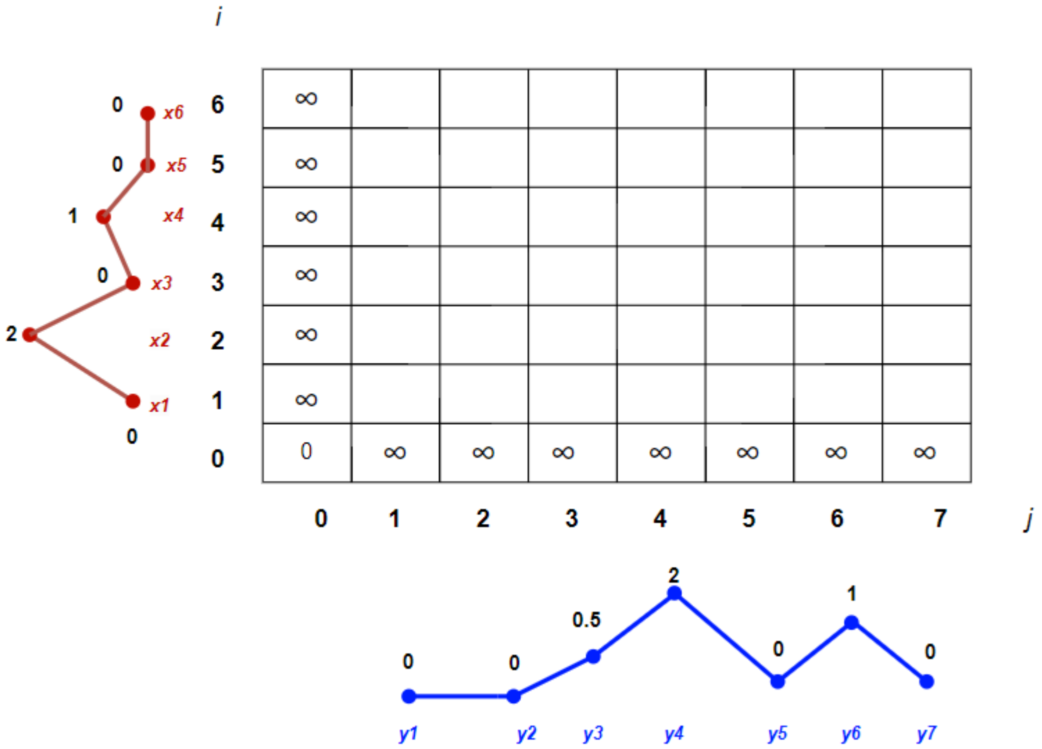


Fig. 17 Cost Matrix Initialization

After completing the initialization phase, the next step is to populate the entire cost matrix by traversing all rows from i=1 to i=N, and for each row, all columns from j=1 to j=M. By convention, this process typically begins at cell D (i=1, j=1).

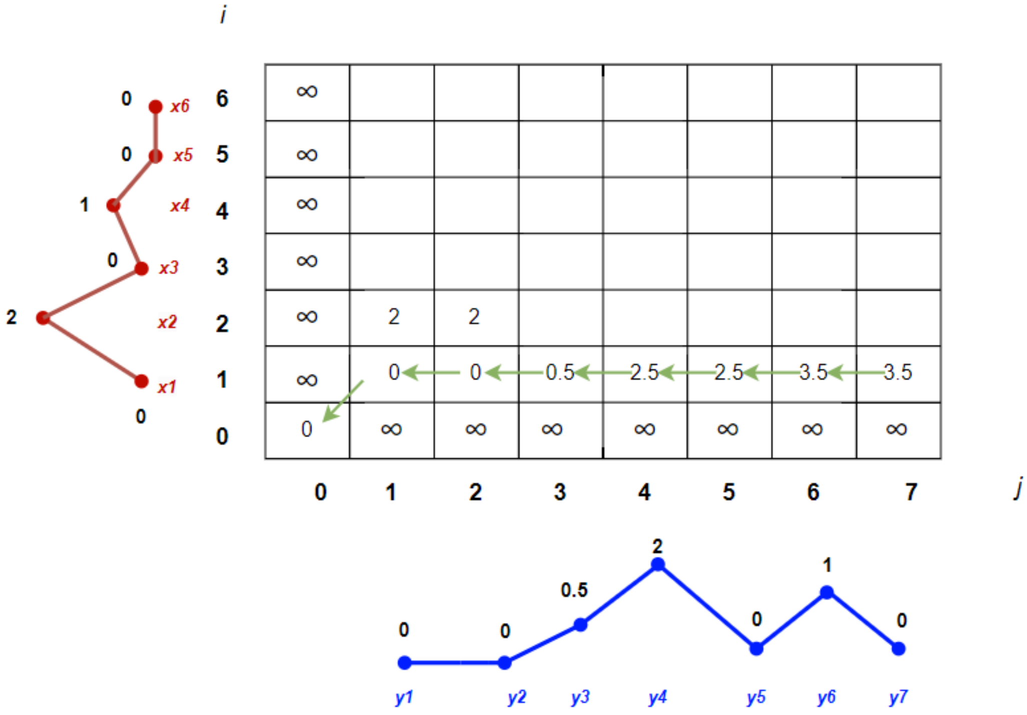
For each cell is calculated:

```
for i = 0 to N:
        for j = 1 to M:
                $D_{i,j}$ = d ($x_i$ , $y_i$) + min {($D_{i-1\,,j-1}$), ($D_{i-1\,,j}$), ($D_{i\,,j-1}$)}
```

- If the minimum is ($D_{i-1\,,j-1}$) → there is a *match*;
- If the minimum is ($D_{i-1\,,j}$) → there is an *insertion*;
- If the minimum is ($D_{i\,,j-1}$) → there is a *deletion*;

Where d (x_i, y_i) is the Euclidian distance useful for calculating the distance between two points x_i and y_i that corresponds to the $|x_i$ - $y_i|$.

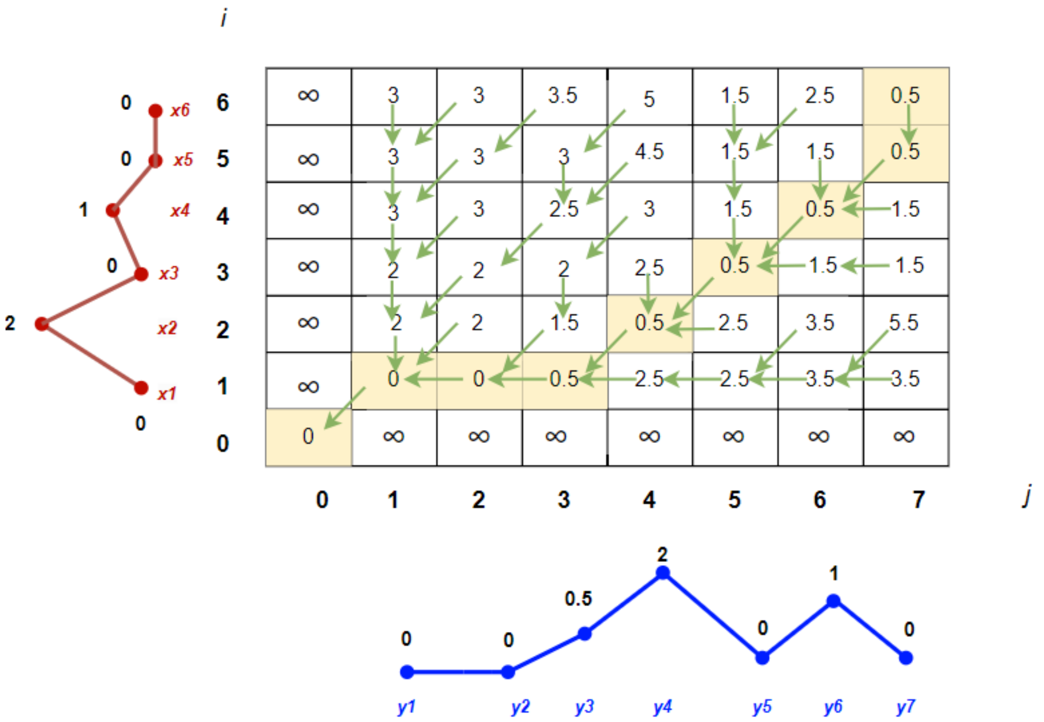
As can be observed, in some cells, the computed value is 0. This indicates that aligning these points with each other incurs no cost.



i								
6	∞							
5	∞							
4	∞							
3	∞							
2	∞	2	2					
1	∞	0	0	0.5	2.5	2.5	3.5	3.5
0	0	∞	∞	∞	∞	∞	∞	∞
	0	1	2	3	4	5	6	7 j

Fig. 18 Cost Matrix with some computed values

After computing all the cells of the cost matrix, the next step is to trace the optimal path from D (N, M) back to D (0,0). The algorithm tells us how to align different signals and it does so by finding the minimum cost path thanks to the cost matrix. It tells us what the most convenient way is to align these 2 signals.



i								
6	∞	3	3	3.5	5	1.5	2.5	0.5
5	∞	3	3	3	4.5	1.5	1.5	0.5
4	∞	3	3	2.5	3	1.5	0.5	1.5
3	∞	2	2	2	2.5	0.5	1.5	1.5
2	∞	2	2	1.5	0.5	2.5	3.5	5.5
1	∞	0	0	0.5	2.5	2.5	3.5	3.5
0	0	∞	∞	∞	∞	∞	∞	∞
	0	1	2	3	4	5	6	7

Fig. 19 The minimum cost path

The result of aligning the x and y sequences by following the optimal cost path is as showed in the Fig.20. As can be noticed, according to the corresponding minimum associated with each element of the cell in the optimal path, there is either a match between two points of the two sequences, and insertion, or a deletion.

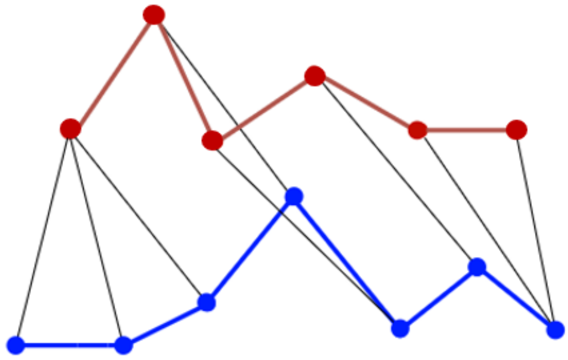


Fig. 20 The result of the alignment

Additional information that DTW algorithm gives is the overall cost of aligning these 2 sequences. This value is equal to:

D = $\sum_i d\ (x_i\ , y_i)$

It represents the computation for calculating the total dynamic time warping distance path cost.

[15]

6.1.2 FastDTW constraints

FastDTW has important constraints that must be adhered to during the process of aligning time series data. These constraints, including the boundary condition, the monotonicity condition, and the step size condition, ensure that the alignment path is coherent and meaningful holding the sequential and temporal nature of the data. By meeting these constraints, FastDTW is able to provide accurate and reliable results in aligning time series, while contributing to the success of a wide range of applications in temporal data analysis.

- **Boundary Condition**: this condition guarantees that the alignment path generated by FastDTW includes both the beginning and the end of the time series. In other words, the alignment must cover the entire duration of both time series under comparison. It is important because it ensures that important information from the beginning or end of the time series is not overlooked during the alignment process. To sum up, the path must include beginning and end.
- **Monotonicity Condition**: fastDTW requires that the alignment path maintain a monotonous progression, avoiding sudden jumps or reversals during the alignment. Essentially, the path should transition smoothly from point to point, following the natural progression of time. This constraint preserves the temporal order of the data points and ensures that the alignment remains consistent and interpretable. To sum up, the path must not have any jumps.

- **Step Size Condition**: This condition stipulates that the alignment path created by FastDTW cannot go backward in time, it must always proceed forward. In other words, the alignment can only progress toward adjacent points in the time series, avoiding abrupt shifts or discontinuities in the alignment. By adhering to this constraint, FastDTW ensures that the alignment accurately reflects the temporal relationships between data points without introducing unrealistic distortions. In summary, the path cannot go back in time.

These constraints collectively ensure that FastDTW produces meaningful and interpretable alignments of time series data, making it a valuable tool for various applications in data analysis and pattern recognition.

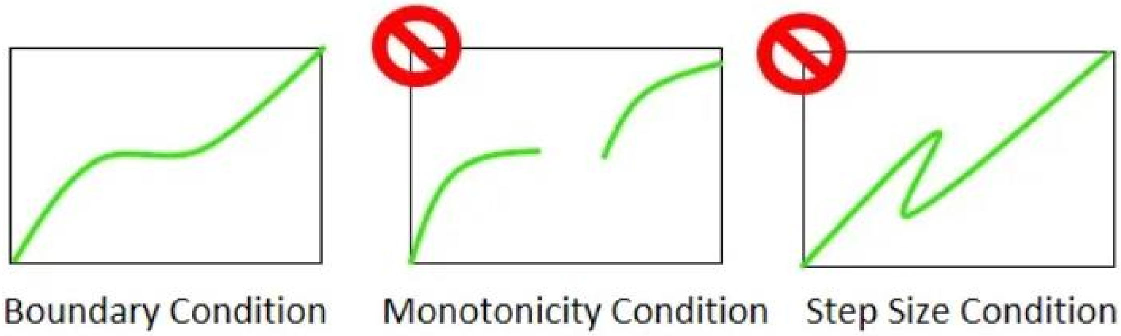
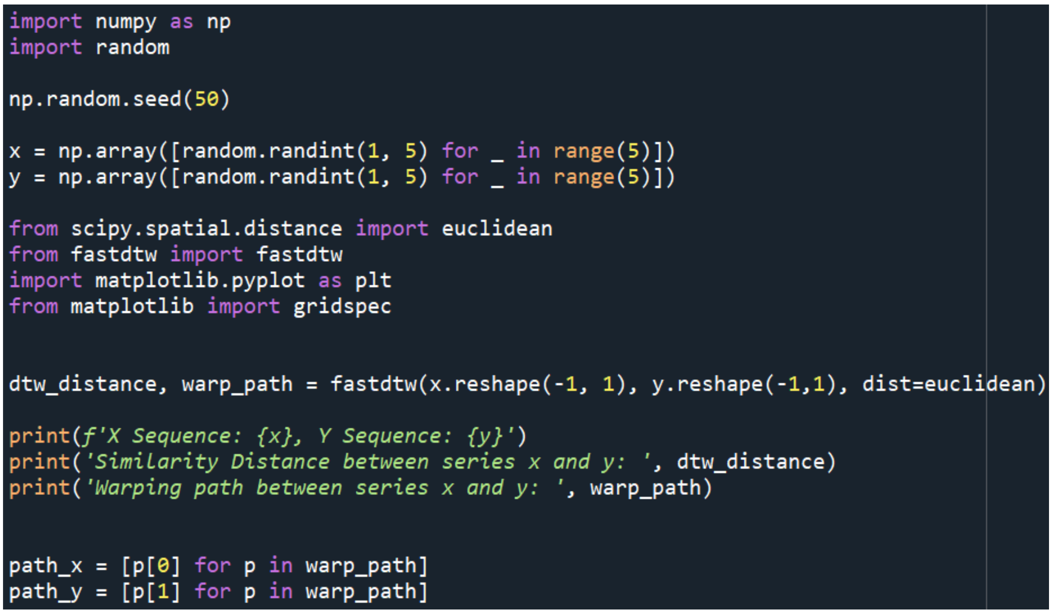


Fig. 21 The constraints of the Fast DTW by https://medium.com/trusted-data-science-haleon/fastdtw-in-action-optimizing-manufacturing-operations-c07f3cc5023c

6.1.3 Implementation

In this paragraph, the implementation of the Dynamic Time Warping algorithm is showed using Spyder IDE. The aim is to present the key steps and logic involved in achieving the desired functionality.

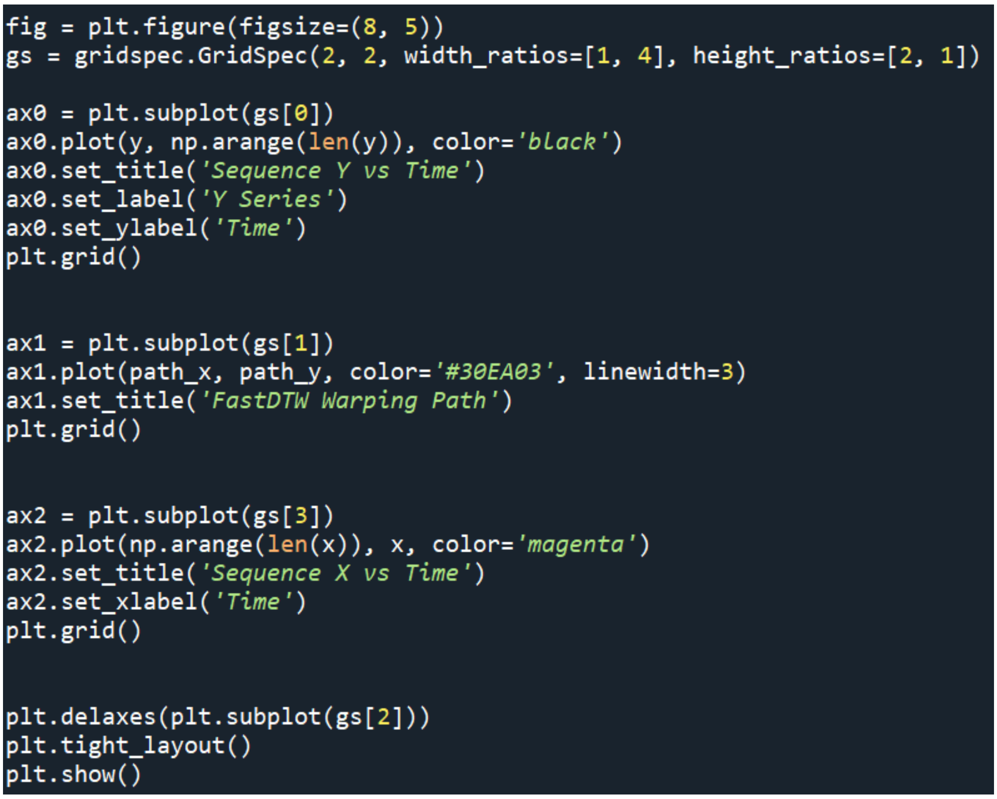
In the first section of the code, all needed libraries are imported such as numpy and random, and then a seed is set in order to allow the reproducibility of the random number sequences. Next, 2 arrays containing random numbers in the range 1 and 5 are generated. Using the FastDTW package, the distance between the 2 sequences is computed and the optimal path that minimize the alignment cost is found.



```
import numpy as np
import random

np.random.seed(50)

x = np.array([random.randint(1, 5) for _ in range(5)])
y = np.array([random.randint(1, 5) for _ in range(5)])

from scipy.spatial.distance import euclidean
from fastdtw import fastdtw
import matplotlib.pyplot as plt
from matplotlib import gridspec


dtw_distance, warp_path = fastdtw(x.reshape(-1, 1), y.reshape(-1,1), dist=euclidean)

print(f'X Sequence: {x}, Y Sequence: {y}')
print('Similarity Distance between series x and y: ', dtw_distance)
print('Warping path between series x and y: ', warp_path)


path_x = [p[0] for p in warp_path]
path_y = [p[1] for p in warp_path]
```

Fig. 22 First Section of DTW implementation

In the second part of the implementation, commands are executed to plot the two sequences and the optimal path. This includes aesthetic characteristics of the sequences and title of the graphs for an immediate visual understanding of the FastDTW. [13]



```
fig = plt.figure(figsize=(8, 5))
gs = gridspec.GridSpec(2, 2, width_ratios=[1, 4], height_ratios=[2, 1])

ax0 = plt.subplot(gs[0])
ax0.plot(y, np.arange(len(y)), color='black')
ax0.set_title('Sequence Y vs Time')
ax0.set_label('Y Series')
ax0.set_ylabel('Time')
plt.grid()


ax1 = plt.subplot(gs[1])
ax1.plot(path_x, path_y, color='#30EA03', linewidth=3)
ax1.set_title('FastDTW Warping Path')
plt.grid()


ax2 = plt.subplot(gs[3])
ax2.plot(np.arange(len(x)), x, color='magenta')
ax2.set_title('Sequence X vs Time')
ax2.set_xlabel('Time')
plt.grid()


plt.delaxes(plt.subplot(gs[2]))
plt.tight_layout()
plt.show()
```

Fig. 23 Second Section of DTW implementation

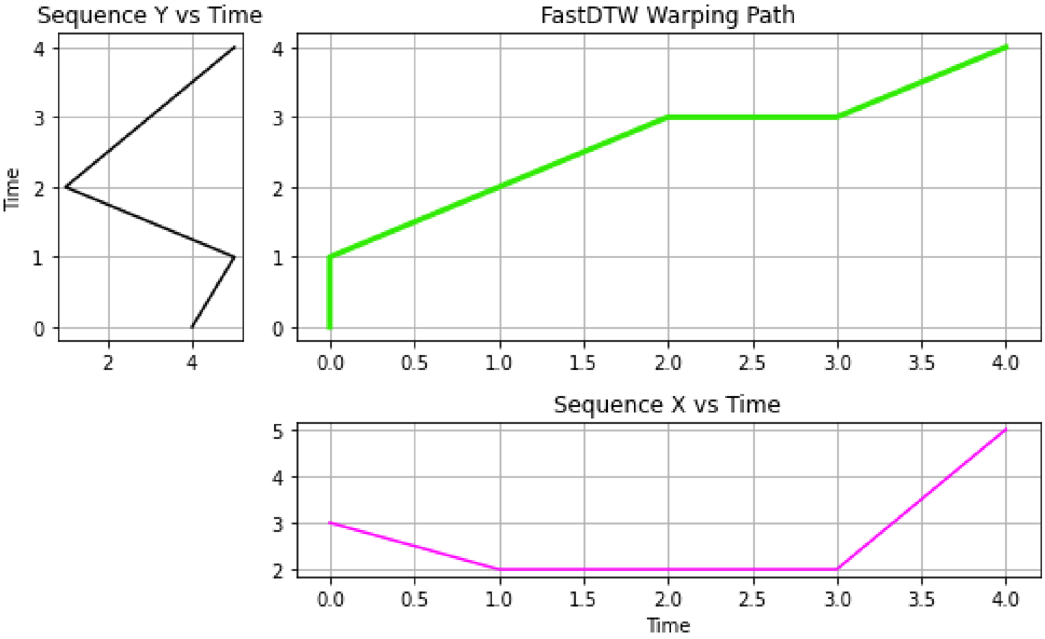
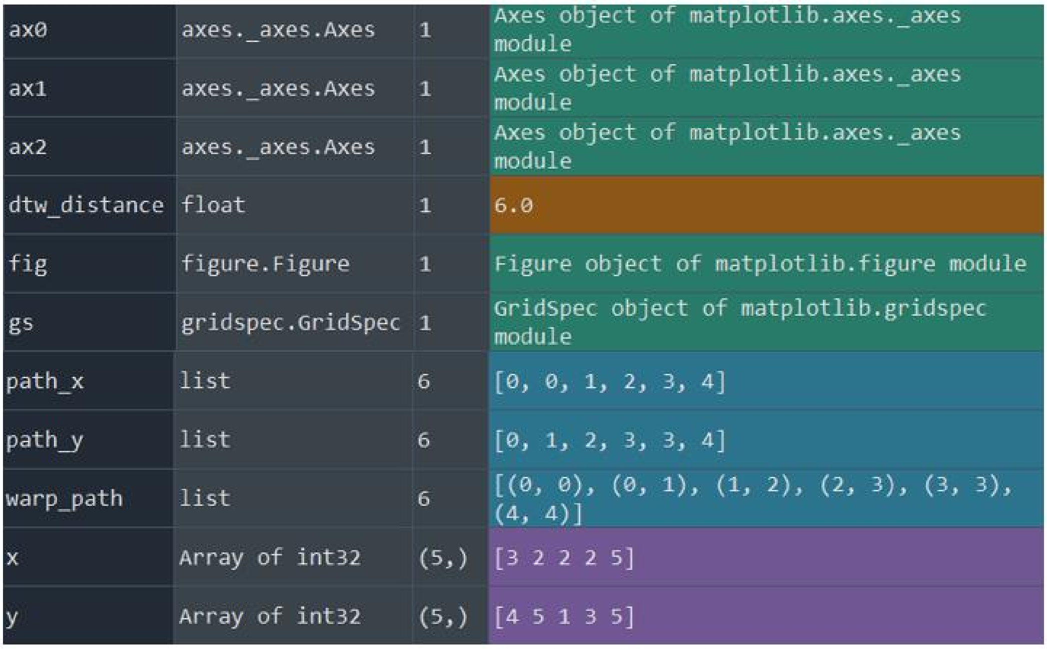


Fig. 24 Plot of X and Y sequences and Warping Path



ax0	axes._axes.Axes	1	Axes object of matplotlib.axes._axes module
ax1	axes._axes.Axes	1	Axes object of matplotlib.axes._axes module
ax2	axes._axes.Axes	1	Axes object of matplotlib.axes._axes module
dtw_distance	float	1	6.0
fig	figure.Figure	1	Figure object of matplotlib.figure module
gs	gridspec.GridSpec	1	GridSpec object of matplotlib.gridspec module
path_x	list	6	[0, 0, 1, 2, 3, 4]
path_y	list	6	[0, 1, 2, 3, 3, 4]
warp_path	list	6	[(0, 0), (0, 1), (1, 2), (2, 3), (3, 3), (4, 4)]
x	Array of int32	(5,)	[3 2 2 2 5]
y	Array of int32	(5,)	[4 5 1 3 5]

Fig. 25 The full set of variables

6.1.4 Application Example in the Manufacturing Context

The Fast Dynamic Time Warping algorithm is used for real-time identification of steps of a toothpaste manufacturing process. By efficiently computing the distance between temporal sequences, even they vary in length or sampling rate, FastDTW enables rapid and accurate analysis of process dynamics. This capability proves invaluable in industrial settings, where timely detection and understanding of production steps are crucial for efficiency and ensuring product quality. Monitoring the toothpaste production process of toothpaste in real-time is crucial due to its complexity and stringent quality standards. By promptly identifying any issues, such as variations in ingredient proportions or inconsistencies in mixing, real-time monitoring enables swift corrective actions to be taken to maintain product safety and efficacy. Given the importance of ensuring the safety of products intended for ingestion, real-time monitoring plays a vital role in guaranteeing that toothpastes production adheres to regulatory requirements and consumer expectations for quality and safety. In a real-world context, the use of FastDTW enables the development of a high-performance data science model that can capture complex relationships and hidden patterns in temporal data, leading to more accurate and informative results.

The production process is characterized by a series of a distinct steps that culminate in the production of a batch of a toothpaste ready for packaging. Each phase of the process is characterized by different features including temperature, pressure, agitator speed, and others, which follow a specific trend during each step. For example, during the mixing phase, the temperature and agitator speed may vary to ensure proper homogenization. Therefore, each step is characterized by a pattern representing ideally how the step should perform to achieve the desired outcomes.

The algorithm feeds on two essential elements: the theoretical patter of all steps discussed above, and the noisy real measurements of how all the steps of the process actually behave in reality. These represents the two inputs (temporal sequences) necessary for the algorithm to be implemented.

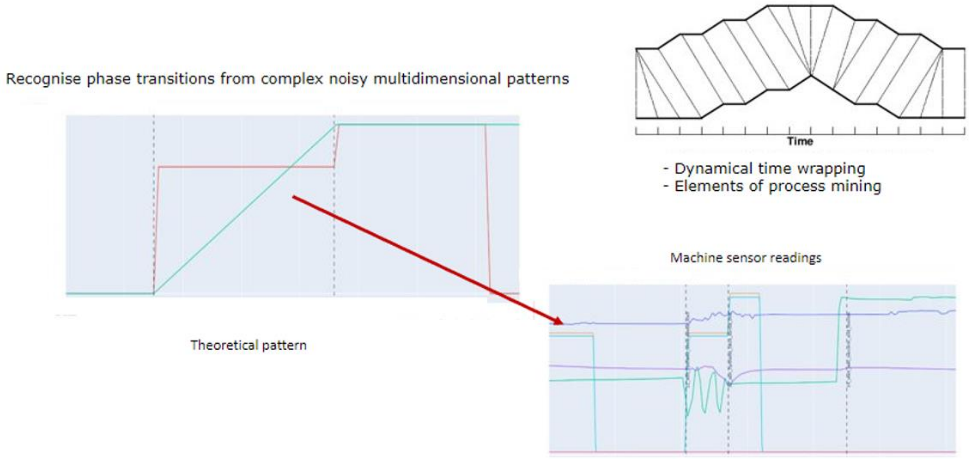


Fig. 26 Example of theoretical pattern and sensor reading of a specific step (y-axis feature values, x-axis time). Axes are omitted to ensure data confidentiality.

Before the algorithm starts working, there are important key aspects that need to be highlighted to clarify how this tool enables the detection of the toothpaste steps.

- **Data Source:** the various real-time data, including temperature, pressure, and other parameters found inside the mixer, are measured by sensors and transmitted from the reference production site's collector to the cloud every 5 seconds. The data science model is seamlessly integrated within the complex digital architecture discussed in the previous chapter, allowing process monitors to track real-time progress through a dashboard.

- **Historical Data Templates:** theoretical models represent one of the most critical phases as they form the basis for phase detection alignment. Based on analyses conducted on large amounts of historical data, reference models for each step of the manufacturing process are deduced. Once these models are tested and validated, a dictionary is created to store how all features change for each step.

- **Distance Threshold:** this threshold is an additional feature that is used in conjunction with the FastDTW calculation that allows the model to identify when step X of the toothpaste production process occurs over time. Without this additional feature, real-time process monitoring would not have been possible. The difficulty that this task requires varies widely, in some cases it is sufficient only to identify what is the threshold that allows the model to identify a global minimum that corresponds to the beginning of step X, in other cases (most) more complex scenarios arise for which it is difficult to find the right value that allows the model to distinguish in all batches local from global minima. This feature of the model will be discussed more in the section called “task 2: Extension of the existing model to a different mixer” where it is represented in detail how to proceed in a practical case in determining the threshold.

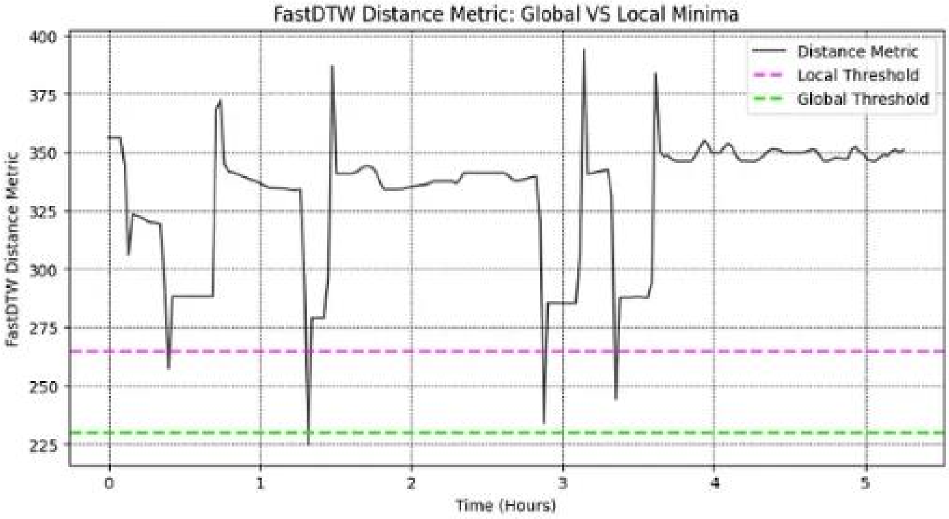


Fig. 27 Distinction between global and local minima by https://medium.com/trusted-data-science-haleon/fastdtw-in-action-optimizing-manufacturing-operations-c07f3cc5023c

- **Model Logic for Step Detection:** since this algorithm is mostly suitable in other fields of applications other than industry, it was necessary for the Data Science team to think of a logic that would allow the steps to always be identified accurately. Consequently, one of the most important features is that the model simultaneously goes looking for both step X and step X+1 so that the model continues to proceed without interruption even if it does not find step X. Another key aspect of the model's logic is that it requires that there be a minimum of X data samples in order for FastDTW to accurately calculate the distance, where X data samples corresponds to the length of the step under analysis.

All these features of the data science model are essential for the ultimate goal to be met. This model allows operators and different production sites to have real-time visibility into processes, enabling rapid adjustments in case of anomalies. If a step X is found to be behaving abnormally, operators are promptly notified allowing effective action to be taken. Once the validity of this model had been analysed and tested, it was immediately thought to adapt it to other products and production sites, that is, to make the model scalable with the aim of adapting the benefits of this logic to many other products and production sites under Haleon's control.

7 Extension of the control system on different site: Oak Hill (NY)

The Data Science team initiated a project to develop an innovative data science model, originally configured to a specific manufacturing site in Maidenhead (UK) and designed to optimize a particular non-aqueous manufacturing process. As the project progressed, it became known that the model's benefits could be greatly expanded by extending its capabilities to other products and production sites. Taking cognizance of this optimization opportunity, a strategic decision was made to create a data science model that would be scalable.

During my research period of the thesis project in Haleon UK, I performed tasks related to the expansion model initiative, bringing my expertise for scaling the model to incorporate Oak Hill (NY) manufacture site. Through collaborative efforts and comprehensive analysis, a model to accommodate the distinctive variables and requirements of the additional site was successfully configured, thereby ensuring the overall effectiveness of the data science model. This paragraph will explore in depth some key aspects of the data science model and its expansion. It will be considered the main characteristics that made the model effective, and it will be analysed how it was adapted to meet the needs of the Oak Hill production site included in the expansion. Subsequently, the tasks performed by me during the model extension process will be thoroughly examined. Through this detailed overview, it will be highlighted the challenges faced, and how they were resolved to ensure the achievement of the scaling process on multiple fronts.

7.1 Key Aspects for the Model Extension: tools&processes, architecture, production.

The exponential growth of data volumes over the last decade together with technological innovation have made Data Science crucial to understand the meaning of data and driving decision making process. Companies face an unprecedented challenge in extracting value from these huge data resources with this ever-increasing amount of information available. Data Science provides with the methodologies needed to analyse, extract meaningful information from data, and interpreter it in a way that enables companies to identify hidden patterns and relationships that are almost impossible to detect otherwise. Through this in-depth analysis, business decisions are informed and based on empirical evidence and data, leading to better, more focused results. In a world, where data has become a strategic asset, data science has emerged as an essential element for the success and competitiveness of companies in various fields.

Scaling the Data Science Model in industrial context means adapting it to handle a higher volume of data from new products and manufacture sites. In this specific case analysed by me, Haleon UK adds a new manufacturing process by an additional site Oak Hill (NY), so it generates more data to be monitored and analysed. The Data Science Model is modified in order to manage this growth, ensuring that it can continue to provide accurate monitoring and control for the optimization of the toothpaste production process.

To make the investment worthwhile, the company needs not only capable data scientists, but also the *tools, processes, infrastructure and production* to support them. [16]

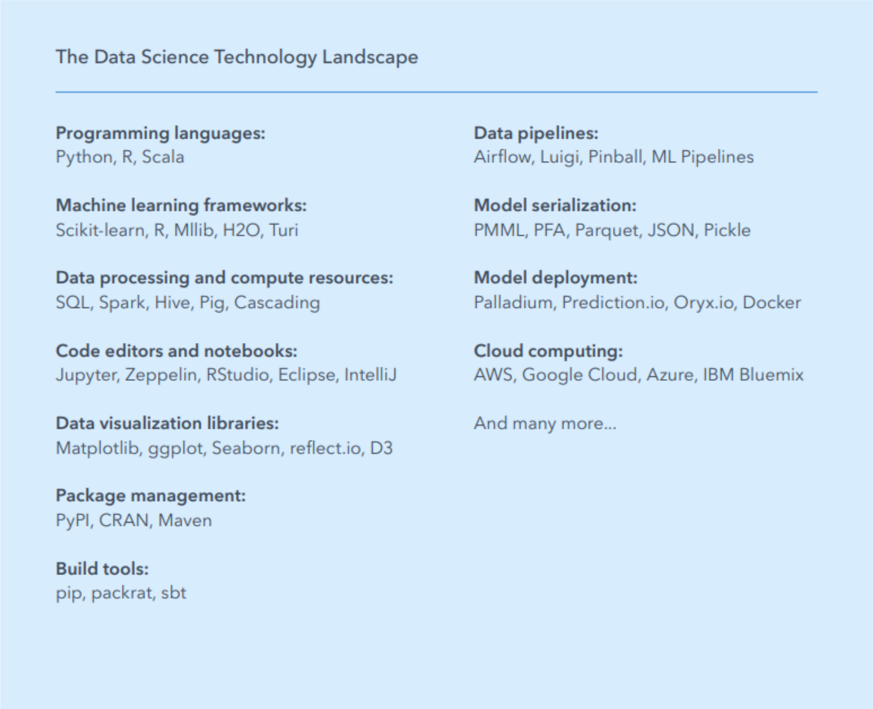
Data Scientists are essential for gaining meaningful insights, but without the right resources and technology environment, their work may be limited. The right tools and processes can make the access to data easier, improve the efficiency in modelling, and encourage collaboration between teams. In addition, a robust infrastructure is crucial for handling large volumes of data reliably and securely, enabling data science to play a central role in business decision-making and allowing a great return on investment.

In the following, the analysis will delve further into the concepts of tools, processes, and infrastructure in the industrial context. These three elements constitute the fundamental pillars that support the entire process of extension of the smart control system for the initiatives of the company. It will be seen in detail how each of these elements contributes to create a favourable environment for analysis and ensures that data science model is effective.

7.1.1 Tools&Processes

When scaling a Data Science Model, a significant challenge to be faced is that data scientists or in general technical people do not have the same skills and preferences on tools. This can be a non-negligible obstacle that could significantly makes slower the process.

Every member team is familiar with specific programming languages, code editors and notebooks…and others analysis tools. Working with a particular platform or tool set may create inefficiencies. Difference in skills and tool preferences requires careful management to ensure that the team can collaborate effectively, share the work and maintain always high the productivity.



The Data Science Technology Landscape

Programming languages:
Python, R, Scala

Machine learning frameworks:
Scikit-learn, R, Mllib, H2O, Turi

Data processing and compute resources:
SQL, Spark, Hive, Pig, Cascading

Code editors and notebooks:
Jupyter, Zeppelin, RStudio, Eclipse, IntelliJ

Data visualization libraries:
Matplotlib, ggplot, Seaborn, reflect.io, D3

Package management:
PyPI, CRAN, Maven

Build tools:
pip, packrat, sbt

Data pipelines:
Airflow, Luigi, Pinball, ML Pipelines

Model serialization:
PMML, PFA, Parquet, JSON, Pickle

Model deployment:
Palladium, Prediction.io, Oryx.io, Docker

Cloud computing:
AWS, Google Cloud, Azure, IBM Bluemix

And many more...

Fig. 28 Different tools by https://www.oracle.com/it/a/ocom/docs/oracle-ds-scaling-data-science.pdf

The work of a data scientist includes several aspects. Among them, the most relevant ones are collection of data, data cleaning, data analysis and finally model building.

Generally, in the initial phase, data scientists must collect data from various sources ***(data collection),*** while ensuring that they are complete. Subsequently, the raw data are cleaned to remove outliers, manage the missing values, and transform them into a desired format ***(data cleaning).*** This process allows to perform an analysis on the basis of solid information.

Data analysis involves the application of methodologies and algorithms to detect useful insights and hidden patterns ***(data analysis).*** The results of the analysis must be clearly interpreted and communicated to the various stakeholders, often through intuitive visualizations and detailed reports in order to support informed decision-making process. At the end, there is the building of the model ***(model building).***

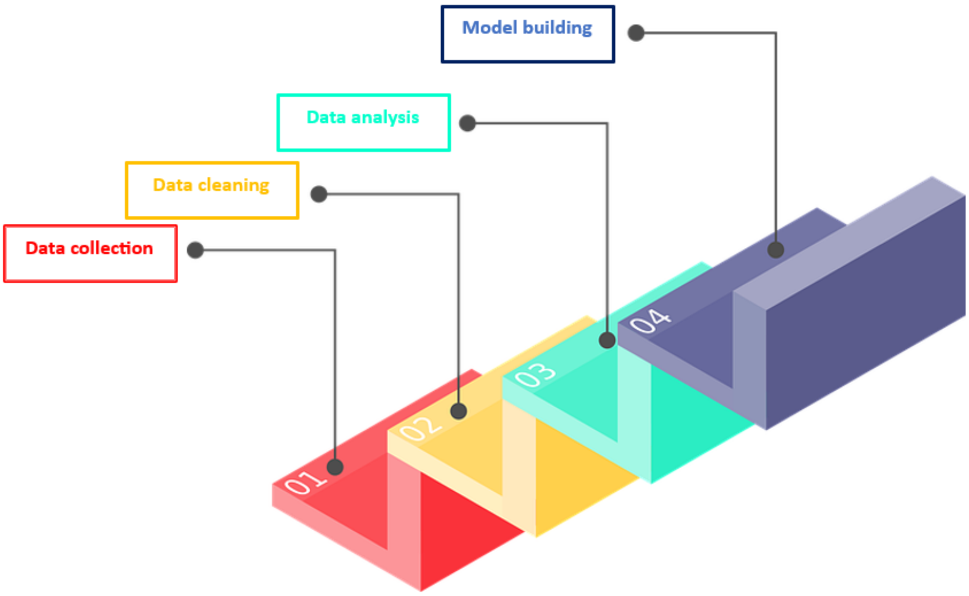


Fig. 29 The most relevant work steps

This process requires advances technical skills and in-depth understanding on how to transform complex data into accessible and usable information.

Throughout this process, it is crucial that these activities are supported by tools and process that enable:

1. Experiment:

To achieve an effective data experimentation, it is highly important to use tools and processes designed to simplify this process. This implies to have systems that allows to collect and manage large amounts of data efficiently. All these characteristics become of fundamental importance especially when the desired result is the extension of an already existing data science model.

It may seem obvious, but it is important to highlight that the larger the amount of data is, the more chances to discover useful information there are. Therefore, the investment in infrastructure and tools that support an efficient data experimentation is crucial for the success of data science initiatives. [16]

2. Create work that is reproducible:

It is crucial to use tools that facilitate the creation of reproducible work and that can be shared among team members. If one member creates a model and stores it locally on its own computer, the other members cannot use it and they will have to build a similar model from the beginning for another task.

A practical example could be Git. It is a Distributed Version Control System (DVSC), a software that allows the source code of a particular project to be effectively managed during its development. Thanks to Git, data scientists can track every change made on the code, compare different versions of the same file, work simultaneously on the same project adding new features without interfering with the work of other team members. [16] [17]

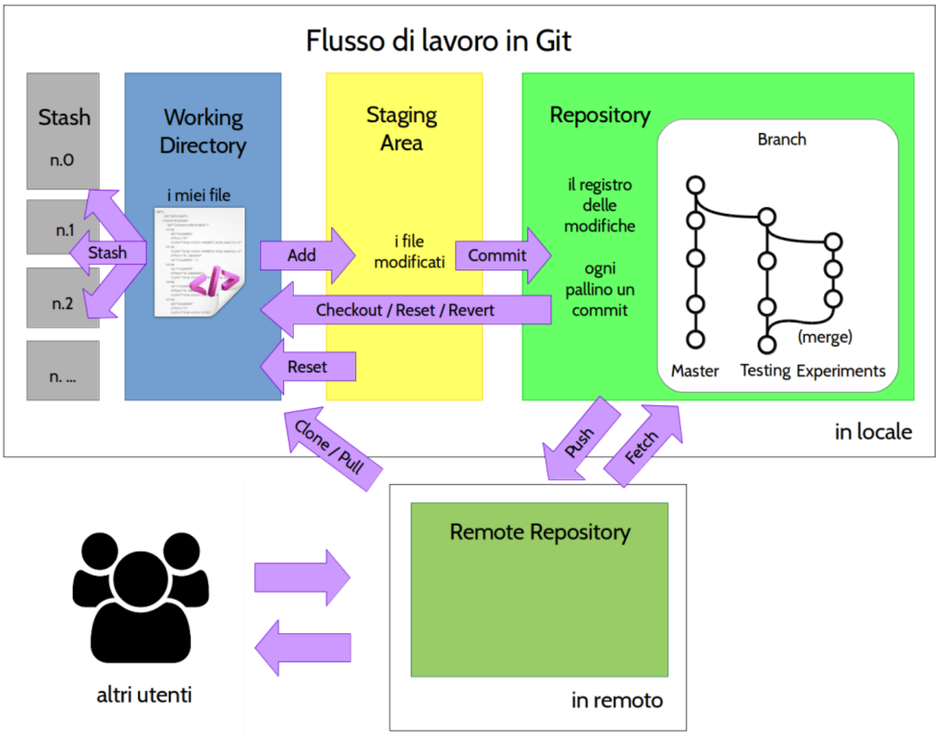


Fig. 30 Git Workflow by http://pierog.it/2018/08/breve-introduzione-a-git/

3. Collaborate and share across teams:

It is crucial to promote collaboration and sharing among the various teams who work on the creation and extension of the control system in order to maximise efficiency and ensure the quality of the end result.

Building sharable reports and enable real-time visualizations by the dashboard may significantly align the knowledge and actions of all members involved. In this context, teams that collaborate for the same goal include data scientists, machine learning engineers, product owners and site managers. Communication and information sharing between these teams allows for synchronisation of efforts and ensures that every aspect of the control system is developed and implemented in a cohesive and focused manner, reacting effectively to project goals and requirements.

To conclude the Tools and Processes paragraph, as stated in the Oracle Article mentioned below, it is unrealistic to impose a specific set of tools on all members of the team. Data Scientists use different programming languages and prefers different integrated development environments (IDEs) such as Jupyter Notebook which I also used during my research period. It is preferable to provide with a platform that can run the code written in any languages within IDEs preferred by each data scientists. This approach allows the team to work flexibility according to their own preferences, while promoting the ability to scale the data science model efficiently. [16]

7.1.2 Infrastructure

It is not only algorithms or data analysis methodologies that are important, but also the entire infrastructure that needs to support these processes. The infrastructure involves hardware, software and environments required to performed data analysis efficiently.

It is essential that the choice of infrastructure is aligned with the specific project goals because a well-designed infrastructure can significantly improve the efficiency and speed of the analysis. An infrastructure adapted to the project goals ensures that scalability requirements are met. Therefore, investing time, costs, and resources in the selection of the most suitable infrastructure not only supports the current needs but also helps to ensure a sustainable future growth for the acquirements of the new requirements. [16]

1. Cloud computing resources

Cloud computing resources are IT services supplied through Internet by cloud computing providers. These resources allow a company to enjoy various services without physical local infrastructure.

One of the key advantageous of cloud computing resources are their scalability and flexibility. They are fundamental characteristics especially for a project that requires scaling up a control system by making the entire infrastructure as flexible as possible. [16]

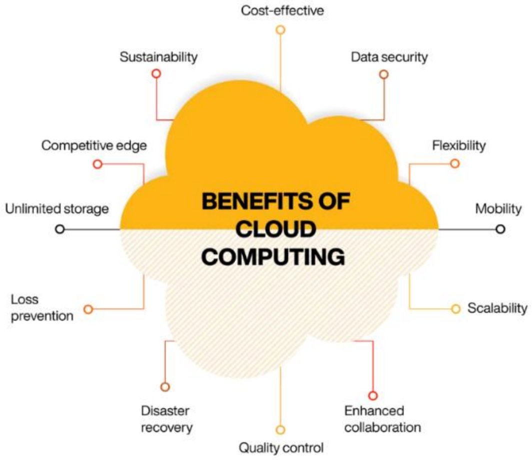


Fig. 31 Advantageous of Cloud Computing by https://www.onlinemanipal.com/blogs/advantages-of-cloud-computing

2. Standardized data science environments

Standardized data science environments offer an effective solution for managing development environments configurations. To set up an environment for each project can be long and complex process. Standardized data science environments allow a creation of “pre-packaged configurations” and it makes possible to save time. [16]

3. Access control

Providing role-based access to data is crucial for the operational efficiency within a company as well as safeguarding security.

Operational efficiency was mentioned because in this way all team members only have access to data related to their work and are not overloaded with irrelevant information. To sum up, a careful management of access control is crucial to ensure a productive work.

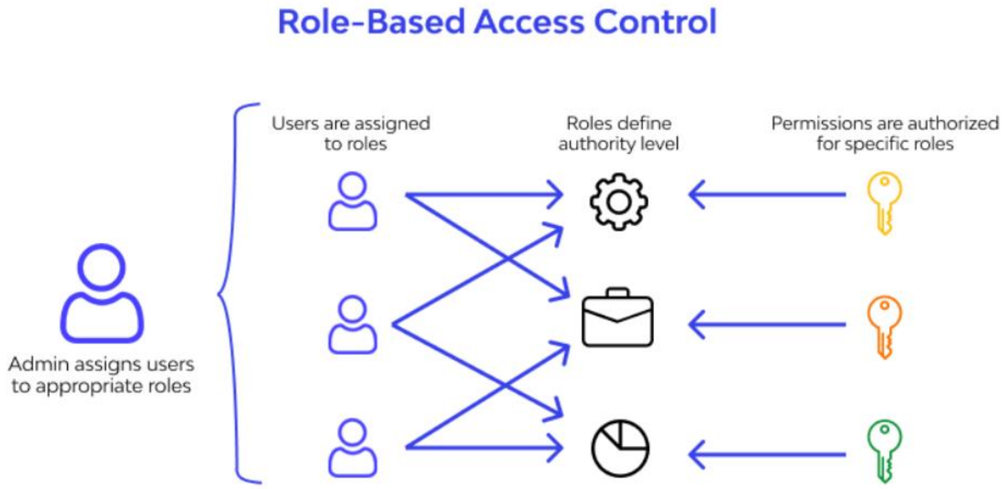

Role-Based Access Control
Users are assigned to roles
Roles define authority level
Permissions are authorized for specific roles
Admin assigns users to appropriate roles

Fig. 32 Role-Based Access Control by https://www.wallarm.com/what/what-exactly-is-role-based-access-control-rbac

All these features listed above could provide a solid starting point for operating in a flexible environment that allows for rapid scalability of a control system. The ability to have a suitable infrastructure ensures the agility to adapt quickly towards the new needs.

7.1.3 Production

After we have integrated the new manufacturing toothpaste process into the data science model, is needed to move on the one of the most challenging phases of the whole cycle: putting it into production.

This phase involves not only the testing and validation with the implicated refinement of the model created, but also the integration of the new piece of the work into the existing system allowing the scalable process.

In this step, the collaboration between data scientists, machine learning engineers and site managers is crucial to overcame technical and management challenges allowing the task to be implemented in the shortest possible time.

The following aspects make easier the putting the data science model into production, making it more streamlined and efficient.

1. Automation low-level tasks

Automating low-level tasks could be a strategically advantageous choice so that bigger amount of time can be spent on activities with higher added value. For example, automating the writing of a report after an analysis saves time by making the work more efficient. Nowadays, through advanced algorithms, can be generated accurate and detailed reports allowing team members to concentrate themselves on more complex tasks. This solution not only helps workers, but also allows greater productivity within the company.

With reference to this specific project, a future step could be to automate labelling: the task that allows to analyse the actual batches and save the start and end time results that will serve as a reference in the tasting phase.

2. Continuous improvements of model performance

Continuous monitoring of the model's performance is required in order to be able to detect potential anomalies as soon as possible and fix them. This task takes place in the testing phase where real time data is collected and are measured deviations from the expected data allowing the model to be improved.

Without constant observation, problems can be undetected compromising the accuracy and reliability of the model. In this way, the problem can be addressed and solved.

7.2 My contribution

These tasks allowed me to understand practically how a model is scaled up by integrating the monitoring of an additional manufacture site. Task 1, called

Labelling, consists of visually identification and storing in files where each step in the manufacturing process begins and ends. This enabled me to realise that only an effective process monitoring enables reliable decision making. Reliable data lead to accurate analysis, which creates the basis for correct decisions. Labelling is fundamental for task 2 which consists of conducting all the necessary steps for the integration process where the Fast Dynamic Time Warping comes into play with all its characteristics. These activities required the use of variable tools allowing to acquire hard skills. Among the most frequently used tools were GitHub, Git desktop, Anaconda, Jupyter Notebook. All of them are interconnected tools for software development, especially in data science. GitHub hosts code repositories and its aim is to improve collaboration, while Git Desktop offers a graphical interface to manage handle repositories and synchronise them with GitHub. In addition, Anaconda manages Python environments, including Jupyter Notebook used for both task 1 and task 2.

7.2.1 Labelling: Task 1

To better understand what this task is and why it is relevant for the analysis, it is important to briefly repeat the two inputs that feed into Fast Dynamic Time Warping allowing it to monitor production processes in real time. The two inputs between which the similarity is computed are as follows:

- Manufacturing Pattern Templates;
- Actual feature data (weight, pre-mix weight, temperature, pressure, agitator speed, scraper speed and homogeniser speed) measured by on-site sensors at production sites throughout the batch production run.

For each production process, there are officially approved pdf requirements that determine how features change step by step in toothpaste production with a high level of accuracy. Despite the high accuracy of models that describe how a batch behaves from its inception to its exit from the mixer, the data measured by sensors in reality are noisy for several reasons:

- Complexity of the system
- Measurement Errors
- External interferences
- Intrinsic noise
- Technological limitations

In summary, there are models available that describe how the process should ideally perform, but the real world is inherently subject to variability, which means that measured signals are inevitably noisy compared to real models. The **labelling task** is a visual identification with high reliability of production process steps that is captured and stored in csv files. This task holds a fundamental importance especially for two reasons. The first is that thanks to the labelling task is possible to observe how a lot of batches perform by creating manufacturing pattern templates based on the information detected during the labelling. This ensures that all templates created for all steps are indeed representative of the actual batches helping the Fast Dynamic Time Warping algorithm to reduce the computational effort. The second reason is that the labelling is not only crucial for the creation of manufacturing pattern template, but also plays a significant role in the testing phase, where measured data from the data science model is compared with the highly reliable data captured visually. For example, during labelling it was recorded that the step X begins at 00.00, while the model data recorded that the same step was detected at 00.20. These 20 minutes discrepancy is analysed to reduce the gap, therefore improving the ability of the model to detect the process performance accurately. Only accurate labelling ensures that the integration process is reliable and optimal.

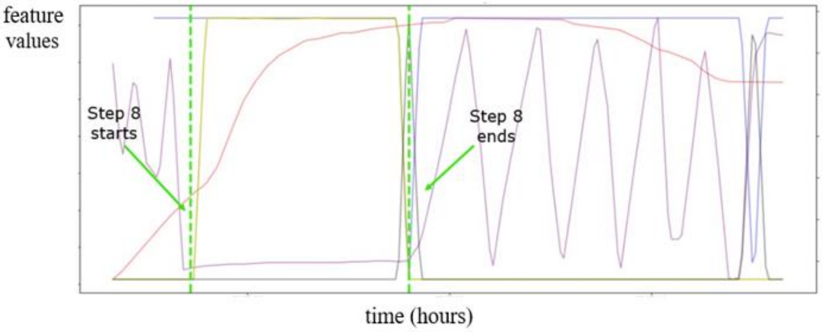


Fig. 33 Visual Identification of a Step

Example

This paragraph is only meant to show how the task was executed and what is the logic behind it. Also, simple cases and more complex cases will be shown for which more in-depth analysis was needed. Because of the data confidentiality clause that the company requires, it will not be possible to show the duration of the whole batch and individual steps, the values associated with the features, and how many steps are involved in producing one batch of toothpaste.

Due to in-depth knowledge of toothpaste manufacturing processes, special characteristics have been identified for each toothpaste manufacturing process that allow for visual proof that the batch we are going to analyse is the correct one.

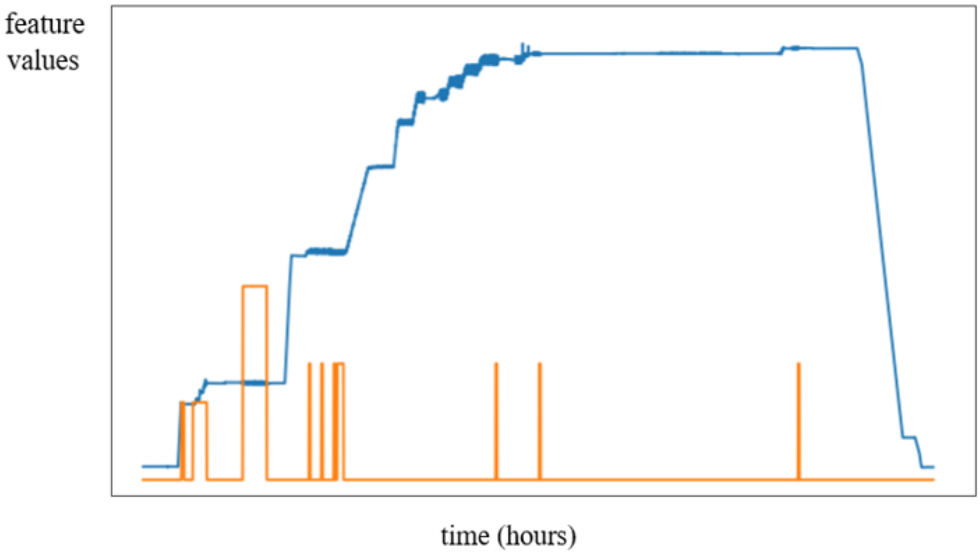


Fig. 34 Visual proof of the whole batch

After making sure that the whole batch is the correct one, is possible to proceed with the study of all the steps of the manufacturing process with the purpose of seeing how the batch performed compared to its ideal behaviour.

Iteratively all steps are analysed. For each step, there are two guide images and the actual data measured by the sensors indicating the actual performance of the batch to be analysed.

A typical case includes:

- **2 reference images**

Fast Dynamic Time Warping Graph shown in the upper left corner of the image and Manufacturing Pattern Templates depicted in the upper right corner.

- **1 batch to be analyzed**

It was shown at the bottom of the image.

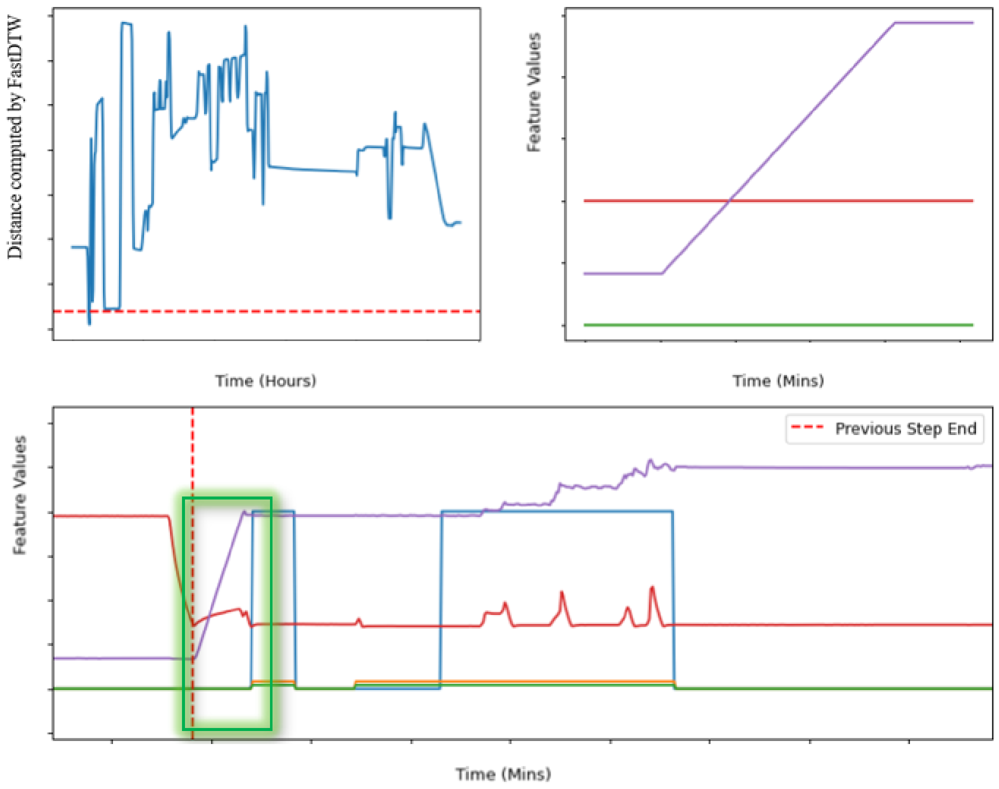


Fig. 35 FastDTW graph, Pattern, and batch to be analysed.

In this example, from the image below, the part of the batch most similar to the Manufacturing Template Pattern shown above will be captured using also the Fast Dynamic Time Warping as a reference. The FastDTW gives more or less accurate indications relative to the beginning of the specific step, which in this case will be between 17.00 and 18.00. In addition, the manufacturing pattern template associated with this step provides information on how the 3 physical features depicted in green, purple and red perform, and the overall duration of the specific step which is about 2.5 minutes. By deriving all this information from the two guide images, it is possible to generate a csv file that keeps the information about this specific step. Start time at about 17.28 p.m. and end time at about 17.32 p.m. The process is repeated for all steps of the toothpaste manufacturing process.

The code structure used in Jupyter Notebook to perform the labelling task is as follows. It helps to get an idea of what are the steps to run it.

- **1st cell:** Authentication tokens.
- **2nd cell:** Import external and local modules.
- **3rd cell:** Plot the whole batch from sql tables.
- **4th cell:** Creation of DTW function taking pattern and batch data from each window.
- **5th cell:** Get Patterns of the specified process at a specific Haleon plant.
- **6th cell:** Plot iteratively DTW graph, pattern of the step X and batch data from each window.
- **7th cell:** Save results in a table.
- **8th cell:** Decision on deleting a row.
- **9th cell:** Save results on a specified path.

Generally, the task is known as **data analysis**. It involves the use of software programmes and tools to examine large datasets in order to extract meaningful information, identify hidden patterns, trends or relationships, and derive new knowledge useful for the organization.

This practice is crucial in many industries where huge amounts of data have to be managed. Data analysis can be conducted using wide range of techniques, including machine learning algorithms, statistics, data mining, data visualization and more.

The end result of data analysis is often the generation of insights that can be used to make informed decisions and identify business opportunities in order to improve the overall performance of the company. It is a powerful tool for innovation and continuous improvement.

7.2.2 Extension of the model to a different mixer: Task 2

Having profoundly achieved a full understanding of what labelling is and the importance of its use for the proper functioning of the data science model, task number 2 is more complex. From a broad point of view, it concerns the extension of the model for including not only Maidenhead (UK) but also from Oak Hill site (US).

At this point it is crucial to emphasize again, as it has already been mentioned in chapter 3 that the model was initially developed to a specific process located in Maidenhead (UK) with the aim of extending the model at a later stage.

Under the control of the London-based Data Science Team I took part in, there are the following manufacture sites whose characteristics are:

- Maidenhead site (UK) where mainly non-aqueous products are produced.
- Oak Hill site (US) where aqueous products are predominant.

More in detail, the task 2 is about the scaling up of the model for the purpose of monitoring an additional non-aqueous toothpaste produced at Oak Hill site with a different type of mixer.

From here on throughout the paragraph will be explained all the pitch of work necessary to configure the smart control system to also monitor the new type of mixer in Oak Hill with the aim to identify all steps in the manufacturing process in near-real time. This represents the pivotal goal when talking about the scaling process for integrating multiple products or multiple Haleon sites.

In order to explain the current status, it is important to highlight that at the beginning of the task 2 assigned to me, the monitoring model is already extended to further production processes in Maidenhead (UK), and it is already adapted for the Oak Hill site but only for a production process working with a specific mixer (type K). The company aims to monitor all typologies of mixers used for toothpaste production that are present in Oak Hill.

What is the action plan?

Below, only the key points are shown to summarise all the work steps. Gradually, each of these points will be examined in detail, while ensuring a comprehensive understanding of the entire performance of the task 2.

1. Clear understanding of the differences about mixers.
2. Creation of manufacturing pattern templates associated with a specific manufacturing process from Oak Hill (NY) with a different mixer called in the following explanation mixer 2.
3. Writing a script for identified step specifications.
4. Make sure the patters are correct.
5. Threshold Setting
6. Testing
7. Implementation

1. Clear understanding of the differences about mixers.

The first thing to do is to gain a clear understanding of the differences about the 2 types of mixers. This foundational knowledge is crucial because every mixer is designed for a specific application and offer unique characteristics. In-depth analysis about them is strictly important to ensure that what happens during the process performed in the mixer 2 at Oak Hill is always comprehended and justified enabling informed decision-making.

Method

The analysis was conducted by collecting information from the two Product Owners of the two manufacture sites. One of the most relevant aspects is that the two mixers have different sizes. This dimensional difference is significant because it can affect the operational and functional aspects of the mixers themselves. A larger mixer might have a higher material handling capacity leading to weight measurements that are not expected to us. In addition to that, the analysis disclosed a difference in the number of measured features. In mixer 1, there are scraper and agitator speeds that are strongly distinguishable, whereas in the mixer

2, these 2 features are combined *(scraper&agitator speed).* This inherently means that the clear distinction between the 2 features in the mixer 1 allows for a more precise monitoring which lead to better optimization process. In contrast, the combination of features in mixer 2 may simplify the creation of manufacturing pattern templates but at the cost of reducing accuracy. This trade-off can impact the overall efficiency and quality of the system depending on the specific application and requirements.

2. Creation of Manufacturing Pattern Templates

The Manufacturing Pattern Templates of each step of the production process are repeatable and structured models which describe how each step should perform. More in detail, every pattern specifies how all features perform over time, ensuring each step is performed consistently and efficiently. These templates provide an ideal model, helping to meet the stringent standards of quality and productivity. By following these patterns, it is possible to ensure that the results meet expectations.

In this manufacturing context, these pattern templates are of fundamentals importance because they ensure that the Fast Dynamic Time Warping algorithm accurately monitor the entire production process being one of the inputs that goes into this algorithm.

The goal of this piece of work is to provide the pattern templates of each step of the non-Aqueous process produced at Oak Hill within a new type of mixer different from the others previously used. A key aspect of this task is that the non-Aqueous Manufacturing Process Performance is already familiar, as it is a process that is currently being monitored at Maidenhead (UK). The only difference is that it is produced using a different mixer with different characteristics leading to treat it as a separate process. Therefore, although the overall process keeps similar, the step specifications of each step require careful considerations and adjustments to ensure an optimal performance and quality.

One of the key advantageous is that a substantial know-how of the process performance is already possessed allowing me to use the templates from the Maidenhead site of mixer 1 as a reference point, facilitating the comprehensive analysis of the difference between the 2 sites or, it would be better to say between the 2 mixers. Leveraging this existing knowledge can streamline the efforts and reduce the time required to setup it.

However, a significant drawback is the lack of complete understanding and familiarity with the new mixer. This unfamiliarity could lead to unforeseen challenges and potential inefficiencies as the process is integrated in the monitoring plan. The new mixer may need additional testing and troubleshooting potentially delaying the whole operational efficiency. This feature led me to predict a longer testing time compared to other processes. Consequently, I planned a testing duration twice as long as the others, to ensure any issues can be properly identified and addressed. This approach aims to minimize risks and guarantee that the integration of the new mixer occurs without compromising the overall operational efficiency.

Method

The method used was established based on the objective to be achieved. By aligning the approach with the desired outcome, it is allowed to focus the efforts on the most critical aspects, maximizing the chances of success and ensuring that the results meet the specified goal.

The identification of the manufacturing pattern templates of Non-Aqueous process from mixer 2 was achieved through a continuous comparison with patterns from mixer 1 which had already been detected by the data science model. The analysis was conducted by examining the behavior of each step in the manufacturing toothpaste process across 5 batches originating from the mixer 2. This comprehensive study involves a detailed investigation into each phase ensuring that every feature was meticulously evaluated. By making an analysis on 5 different batches, it was possible provide a robust dataset that allowed for a deeper understanding of how the new mixer influenced the overall production process. This approach allowed the identification of patterns, anomalies, and areas for improvements.

Mixer 1	Weight	Pressure	Scraper& Agitator Speed	Homogeniser Speed	Temperature	Duration Time
$\boldsymbol{Step_n}$	W_n	P_n	$S\&A_n$	H_n	T_n	t_n

Table 1 Features of a generic step from the mixer 1.

The table 1 refers to Mixer 1, it contains data already well-known from the Data Science Team. This table consolidates all the information available for each step of the process. It provides any relevant observations of the mixer 1 supporting the analysis in the comparison with the new data measured from the mixer 2. The table shows the most relevant features: weight, pressure, scraper&agitator speed, homogenizer speed, temperature, and duration time of each step.

$\boldsymbol{Step_n}$	***It represents the reference step***
$\boldsymbol{W_n}$	*Weight of the step n*
$\boldsymbol{P_n}$	*Pressure of the step n*
$\boldsymbol{S\&A_n}$	*Scraper&Agitator speed of step n*
$\boldsymbol{H_n}$	*Homogeniser speed of step n*
$\boldsymbol{T_n}$	*Temperature of step n*
$\boldsymbol{t_n}$	*Duration time of step n*

Naturally, there would as many tables as the number of steps in the manufacturing process. For simplicity, only 1 table is depicted. This approach allows for an easier understanding of the method while avoiding redundancy of displaying multiple tables with excessive data.

Mixer 2	Weight	Pressure	Scraper& Agitator Speed	Homogeniser Speed	Temperature	Duration Time
$\boldsymbol{Step_n}$						
$\boldsymbol{Step_{n,1st\ Batch}}$	$W_{n,1}$	$P_{n,1}$	$S\&A_{n,1}$	$H_{n,1}$	$T_{n,1}$	$t_{n,1}$
$\boldsymbol{Step_{n,2nd\ Batch}}$	$W_{n,2}$	$P_{n,2}$	$S\&A_{n,2}$	$H_{n,2}$	$T_{n,2}$	$t_{n,2}$

$\boldsymbol{Step_{n,3rd\ Batch}}$	$W_{n,3}$	$P_{n,3}$	$S\&A_{n,3}$	$H_{n,3}$	$T_{n,3}$	$t_{n,3}$
$\boldsymbol{Step_{n,4st\ Batch}}$	$W_{n,4}$	$P_{n,4}$	$S\&A_{n,4}$	$H_{n,4}$	$T_{n,4}$	$t_{n,4}$
$\boldsymbol{Step_{n,5st\ Batch}}$	$W_{n,5}$	$P_{n,5}$	$S\&A_{n,5}$	$H_{n,5}$	$T_{n,5}$	$t_{n,5}$
Mathematical Mode	$v_{0,W}$	$v_{0,P}$	$v_{0,S\&A}$	$v_{0,H}$	$v_{0,T}$	$v_{0,t}$

Table 2 Data collected of a generic step for 5 batches from Mixer 2

The table 2 is more complex because it represents the data collected for each step based on 5 analysed batches. The primary goal is to set the specifications for each step in the most reliable way. The table serves as a crucial tool allowing to document variations and identify consistent trends within the data from mixer 2. By capturing these real transitions ensures that the patterns were constructed on the basis of empirical evidence and actual performance metrics. This method not only enhances the reliability of the patterns but also provides an effective understanding of the new mixer characteristics.

The result obtained for each step is the following table:

Mixer 2	Weight	Pressure	Scraper&Agitator Speed	Homogenizer Speed	Temperature	Duration Time
$\boldsymbol{Step_n}$	$v_{0,W}$	$v_{0,P}$	$v_{0,S\&A}$	$v_{0,H}$	$v_{0,T}$	$v_{0,t}$

Table 3 The resulting features of Step n from mixer 2

$\boldsymbol{Step_n}$	***It represents the reference step***
$\boldsymbol{v_{0,W}}$	*Median (W1, W2, W3, W4, W5)*
$\boldsymbol{v_{0,P}}$	*Median (P1, P2, P3, P4, P5)*
$\boldsymbol{v_{0,S\&A}}$	*Median (S&A1,S&A2,S&A3,S&A4,S&A5)*
$\boldsymbol{v_{0,H}}$	*Median (H1, H2, H3, H4, H5)*
$\boldsymbol{v_{0,T}}$	*Median (T1, T2, T3, T4, T5)*
$\boldsymbol{v_{0,t}}$	*Median (t1, t2, t3, t4, t5)*

Why the median instead of the mean?

The reason why the median was calculated to identify the final value of each feature instead of the mean is because the mean would not have been representative of any real batch. In data with significant variability or outliers, the mean can be strongly influenced by extreme values, thus distorting the true central tendency.

On the other hand, the median provides a more accurate representation of the typical value of each feature within batches because it effectively captures the parameter based on the value that occurs most frequently in batches. This approach ensures that the calculated parameter is more aligned with what is commonly observed, making it a more reliable and robust measure for our analysis.

Importance for testing

The tables constructed for each step of the manufacture process coming from the mixer 2 containing the several values of each feature resulted to be of crucial importance, especially in the subsequent testing phase. Testing requires adjustment and improving the model by modifying the values ensure optimal operation. Having these tables with different values for 5 batches was crucial in guiding us in adjustment the specifications allowing to use them as a reference in our fine-tuning process. By providing a clear and detailed overview of the various parameters, they enable to take informed decisions. This systematic approach not only improved the accuracy of the adjustments, but also ensured that the best possible performance from the model was constantly achieved while highlighting substantial delays in activity.

3. Writing a script for identified step specifications.

After identifying the step specifications, the next piece of work is to write a script that has as inputs all the values related to this particular process coming from the mixer 2. This task work is necessary to ensure that the data science model works correctly.

The following example is a simplified version of the real script for one of the steps. Obviously, the real script has more complex structure which for reasons of data protection it cannot be disclosed. This example is only intended to show the most relevant values of the script, neglecting the actual structure in which all this is embedded. The complete script contains python libraries, importation of local and external modules, data management functions, procedures, and specific implementation details that are fundamental for the correctly operation of the smart control system. By focusing on these key values, the aim is to highlight the critical elements and their importance in the overall scaling process. This example shown below ensures that data confidentiality is respected, while providing valuable insight into the key components of the process.

For the purpose of showing a documentation example of one of the steps, the real script shows have as inputs the specifications step before the step n occurs, the ones after step n occurs and how long the transition is, the stable (unchanging) time before the step started, and stable (unchanging) time after. In other words, it is possible to deduct that the pattern is a simple linear interpolation between the two, with length of each period specified by *stable_time_before, transition_time* and *stable_time_after.*

A super-simplified example of the features associated with each step is shown below.

$\boldsymbol{Step_n}$

Before_transition = {

"Weigh": $v_{o,W}$

"Pressure": $v_{0,P}$

"Scraper&Agitator Speed": $v_{0,S\&A}$

"Homogenizer Speed": $v_{o,H}$

"Temperature": $v_{0,T}$

}

After_transition = {

"Weigh": $v_{o,W}$

"Pressure": $v_{0,P}$

"Scraper&Agitator Speed": $v_{0,S\&A}$

"Homogenizer Speed": $v_{o,H}$

"Temperature": $v_{0,T}$

}

Stable_time_before: t0

Transition_time: t

Stable_time_after: t1

4. Make sure the patters are correct.

Before moving on to the task of Threshold Setting, it is essential to ensure that the patterns defined for all steps reflect what is expected. The approach used to achieve it is to plot iteratively the manufacturing pattern templates for all steps of the process and obtain a visual confirmation of what is aiming for. This visual proof allows to verify that the manufacturing pattern templates are in line with the expectations and requirements by examining deeply each step and identifying any discrepancies from desired templates.

This task is crucial because it provides a clear and immediate representation of the data making easier to make necessary changes. Ensuring that the manufacturing pattern templates are accurate and representative of wanted results is especially critical for the success of the subsequent threshold setting activity. By taking the time to visually validate each step allows to proceed with confidence, knowing that the foundations are solid, and that the threshold will be set on the basis of reliable and accurate data. This meticulous approach helps to minimize errors and improve the overall effectiveness of the scaling process.

As is possible to image, the step name on which is wanted to perform the analysis is inserted as an input, and the code provides us with the pattern associated with that specific step as output. The following image is an example of a plotted pattern.

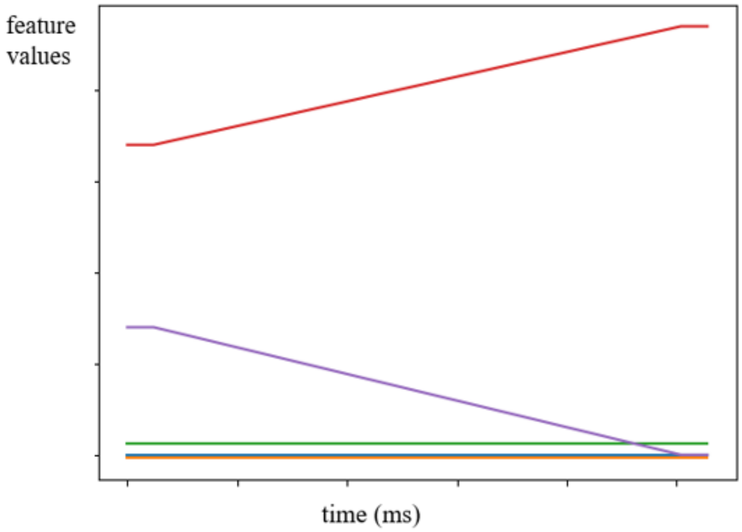


Fig. 36 Example of a plotted pattern

5. Threshold Setting

The distance threshold is an additional feature that the Fast Dynamic Time Warping algorithm uses to precisely detect when the process step occurs. This is achieved by means of continuous comparison between the predefined patterns (the manufacturing pattern templates) and the real-time readings. By constantly monitoring, the Fast Dynamic Time Warping identifies the points where the two sequences are most similar. At these points of high similarity, a global distance is computed, which is then captured thanks to the threshold setting. This threshold allows the model to effectively detect the exact positions within the time series, while ensuring precise temporal alignment and pattern recognition.

The goal of the threshold setting task is the optimization of computational calculations. This helps to focus the analysis only on relevant data ranges, avoiding unnecessary processing on insignificant data. This approach allows essential information to be filtered, reducing the system workload and increasing the speed at which steps of the process are detected. To sum up, threshold serves as a filtering and optimization tool, improving the accuracy and efficiency of monitoring, ensuring that computational calculations are performed in an effective and targeted manner.

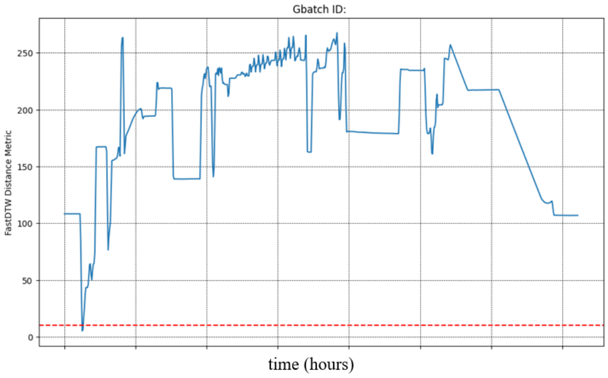
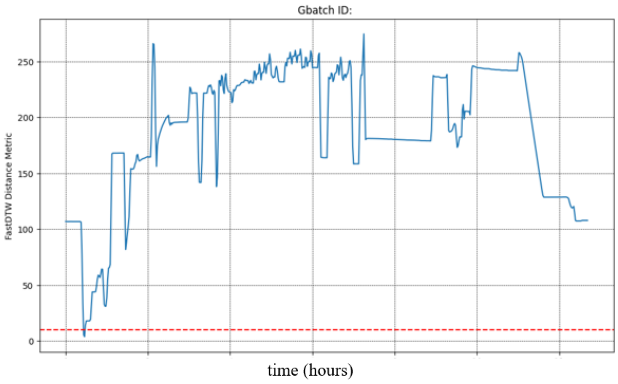
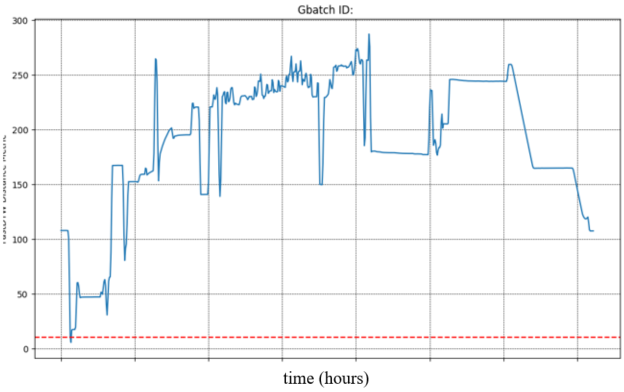
The threshold is set in order to be valid for as many batches as possible. In this specific analysis, 14 batches are considered. This approach ensures that the chosen threshold is flexible and adaptable enough to cover a spread range of scenarios. The desired result is to maximize the overall effectiveness of the threshold, while ensuring that it operates optimally without the need for frequent adjustments. This strategy not only makes easier the identification process, but also decreases potential errors. The ideal situation would be achieving a balance where the threshold is neither too strict not too permissive, providing an optimal solution that meets the different need of all 14 batches under the analysis.

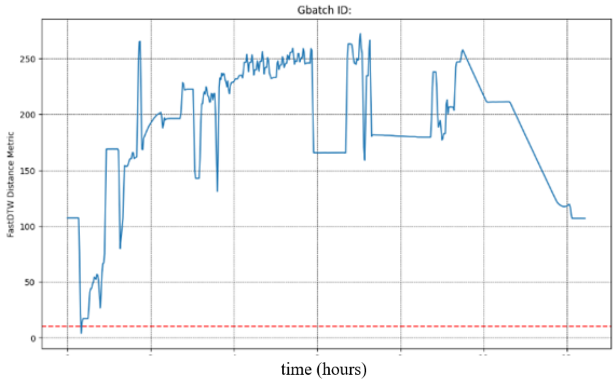
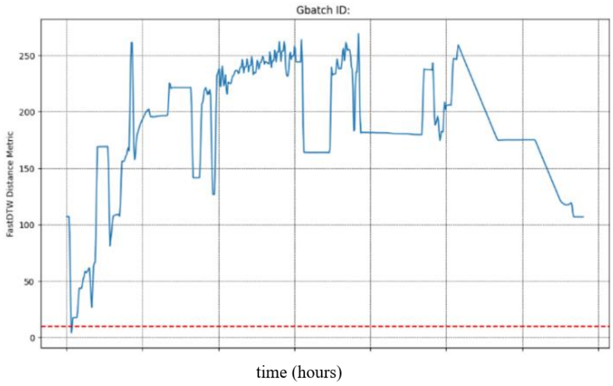
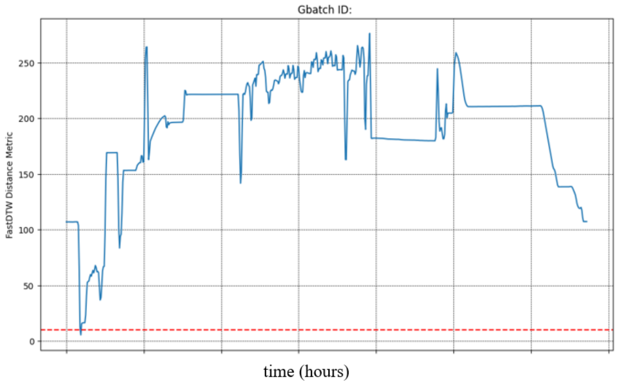
To make the situation clearer, it is emphasized that for each step of the manufacturing process, the threshold is set by analyzing 14 batches. Precise thresholds are set for each step, based on a deeply analysis of these batches, thus ensuring that each step is easily identifiable.

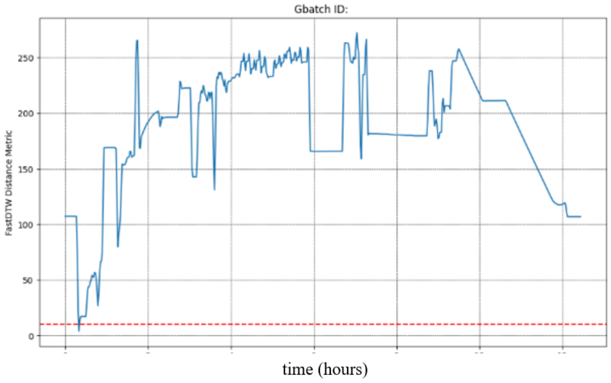
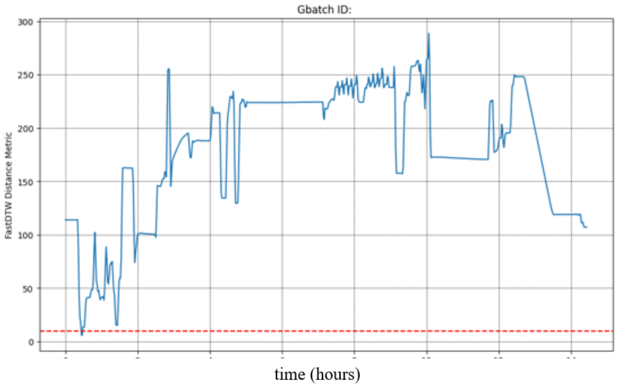
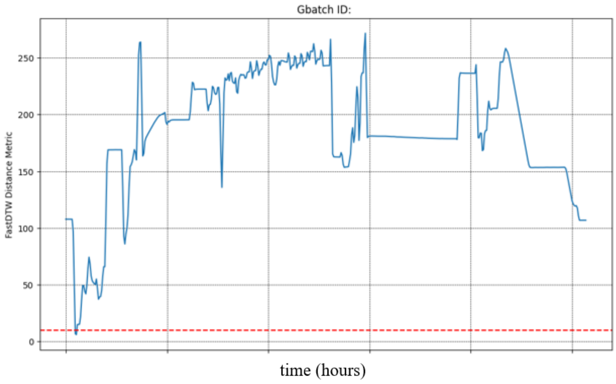
There are steps for which it is immediately easy to define a threshold that fits all the batches involved due to clarity and consistency of the available data. However, there are special cases of steps where threshold setting requires in-depth analysis due to the complexity of the Fast Dynamic Time Warping distance graph. In these cases, detailed analysis is required to ensure that the threshold setting is appropriate for the identification of that specific step.

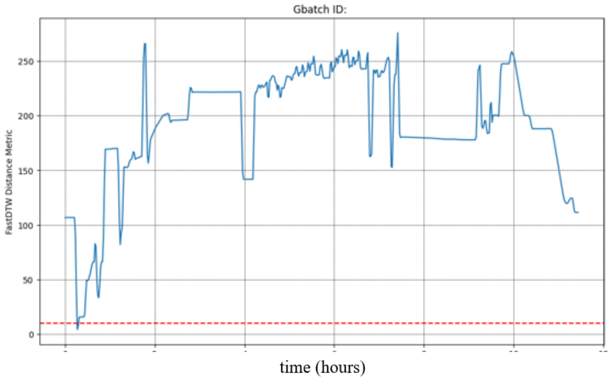
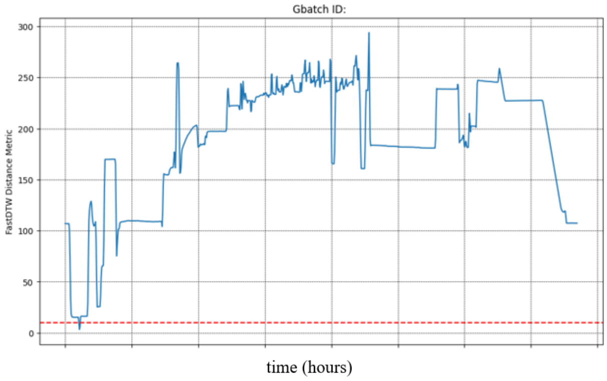
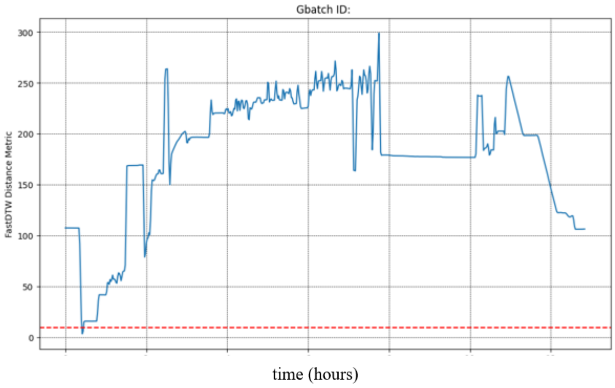
The following is an example of a step for which it is easy to set the threshold: at this stage of the production process, on the first attempt, it is clear that a threshold can be identified that captures the overall minimum for all 14 batches analyses.

The table refers to a specific step of the toothpaste manufacturing process coming from the mixer 2.

ID BATCH	Fast DTW graph
1st Batch	
2nd Batch	
3rd Batch	

4th Batch	
5th Batch	
6th Batch	

7th Batch	
8th Batch	
9th Batch	

10th Batch	
11th Batch	
12th Batch	

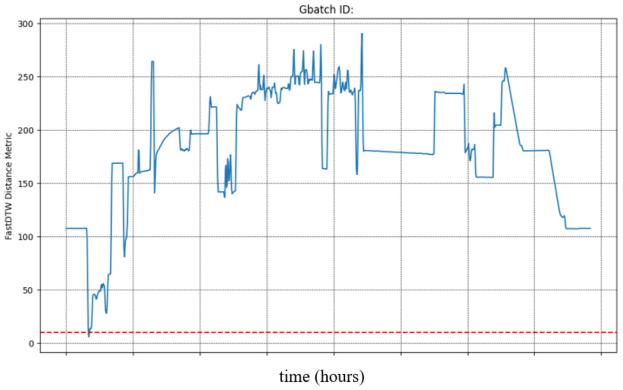
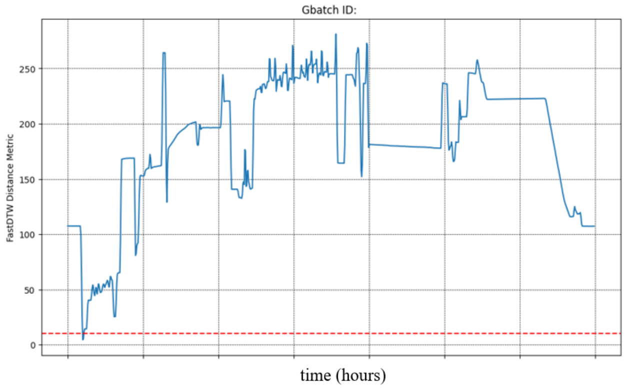
13th Batch	
14th Batch	

Table 4 Example of a FastDTW scenario that makes the threshold setting simple.

Examples of special cases

The following step requires an in-depth analysis due to the complex Fast Dynamic Time Working graph. It is required to analyse in depth the scenario of the *Step X* to have a clear understanding of when effectively the global minimum is. Generally, the graphs show more minima (the "rectangular" and the following "peak").

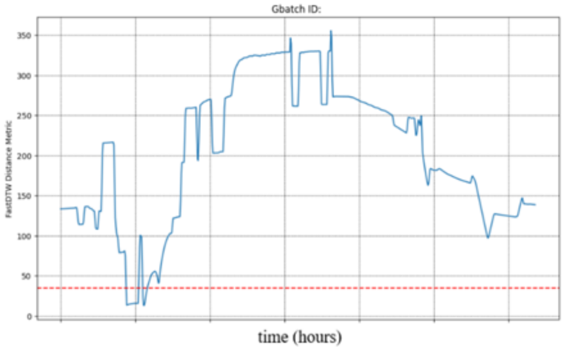


Fig. 37 FastDTW graph without a single minimum

The method that was chosen to follow for the threshold setting is:

- See the labelling results for the start date of step X in order to identify which is the correct minimum in the FastDTW graph.

- Once having identified the correct minimum, you have to make sure that the algorithm detects the correct minimum as the point at which the step occurs. A good way to proceed is to figure out where the algorithm starts working from.

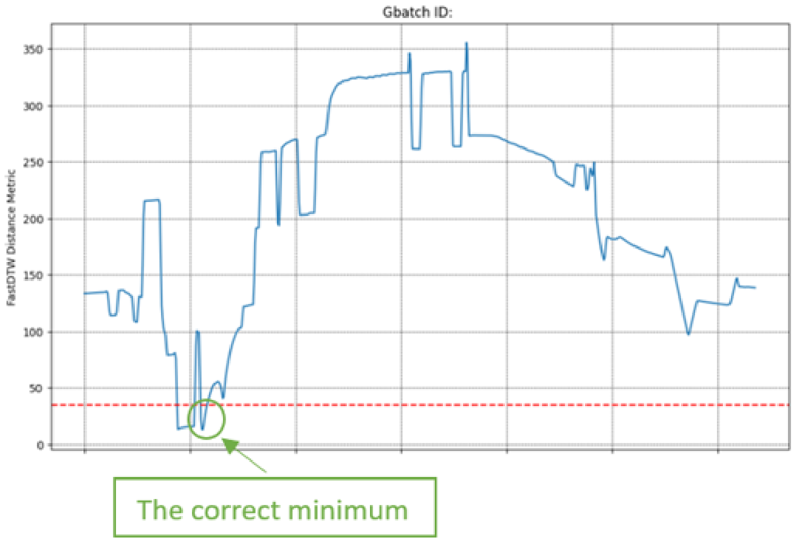


Fig. 38 The correct minimum detected.

Once the correct minimum has been found, it is necessary to ensure that the algorithm detects exactly that minimum and not another.

There are 2 available alternatives:

1. If the previous and the following steps are well identified so that they show a clear global minimum in the FastDTW graph, you can deduct where is the area in which the algorithm works in order to understand which minimum between the two is in that area. This is a quicker method, but it is not always applicable. Below, the

FastDTW graph step X-1 and step X+1 are shown respectively. From their graphs, it is possible to have an idea of where is the area in which the algorithm.

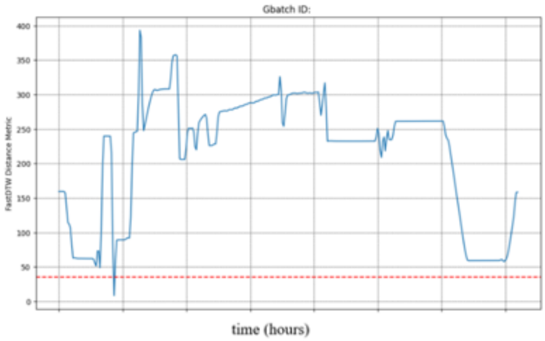


Fig. 39 The clear minimum of the previous step

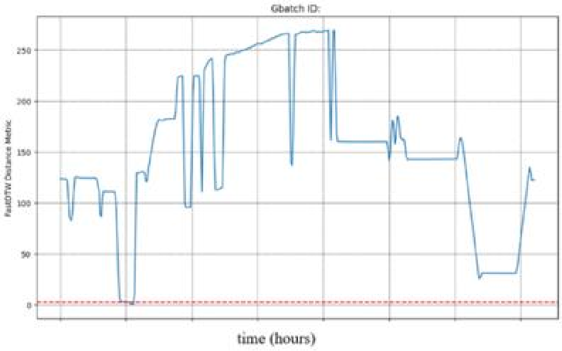


Fig. 40 The clear minimum of the following step

2. Alternatively, you can obtain more information by taking a look at the stored labelling results of the batches under the analysis to make sure from where the algorithm starts working. From the labelling results, it is possible to get exactly information about where the step X-1 end date is. If the end-date of the step X-1 is after the minimum shown by a sort of “rectangle” it will mean that the algorithm starts working just after the “rectangle” ensuring that it identifies the correct minimum. It is now possible to know with certainty what the correct minimum is and there the algorithm starts working from, giving the confirmation that this threshold allows FastDTW to detect this step correctly.

Step X-1 (end date)	*time*
Step X (start date)	*time*

Table 5 Example of the labelling results

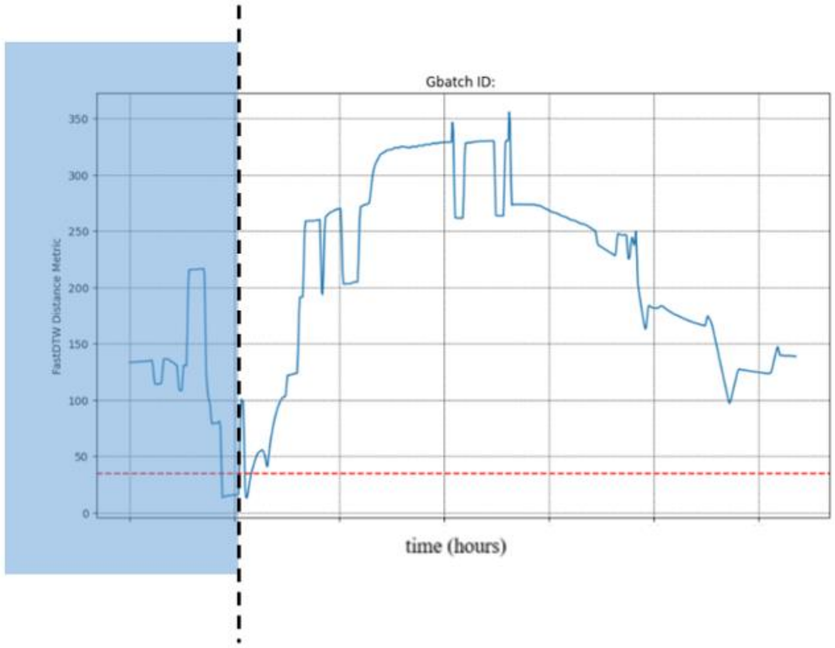


Fig. 41 Visual proof of the FastDTW successfully detecting the minimum.

Now, it is possible to be sure that the first minimum that the algorithm detects is the correct one! With a threshold=38 the algorithm is able to identify the right starting point.

There were also much more complex cases than the last one shown above that required even more evaluations to do to correctly set the threshold. Among the various features measured (pressure, temperature, homogeniser, scraper and agitator speed, etc…), some are less controlled than others. As a result of this variability, the minima might not correspond to the exact point where the step occurs. However, the higher the number of features is, the more it is promoted the specificity that allows to detect these steps. Consequently, to set the threshold correctly, it is essential to find a good trade-off between variability and specificity.

In the following table, an example is shown considering only 4 batches out of 14 analysed for simplicity, as the reasoning for the remaining batches is identical. The second column (Fast DTW graph) shows graphs where it is challenging to set a threshold suitable for all 4 batches due to the complexity of the graphs generated by the algorithm. In the last column (Fast DTW graph without the most variable feature), it is possible to see how the scenario significantly improves, making it easier to set a threshold after removing the most variable feature depicted in red in the following image.

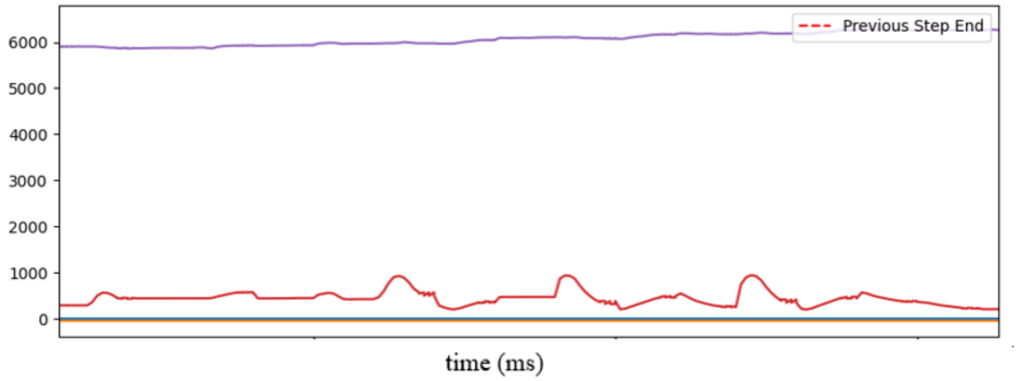
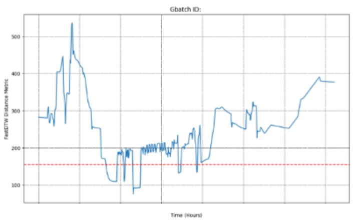
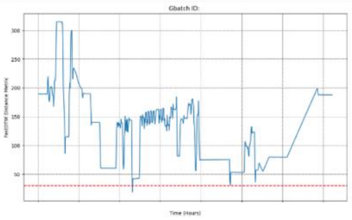


Fig. 42 The most variable feature

ID Batch	Fast DTW graph	Fast DTW graph without the most variable feature
1st Batch		

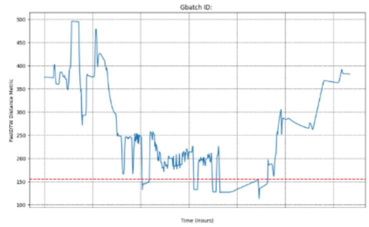
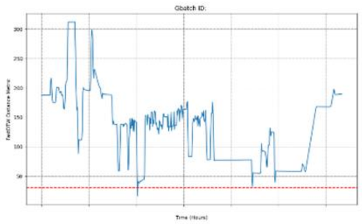
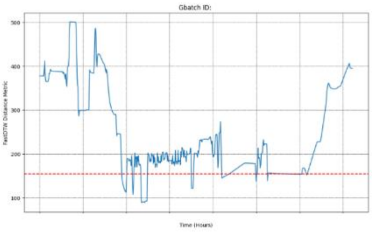
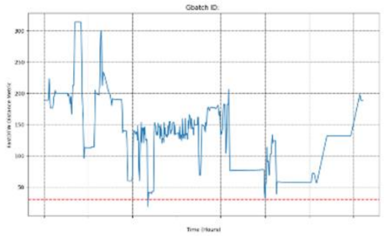
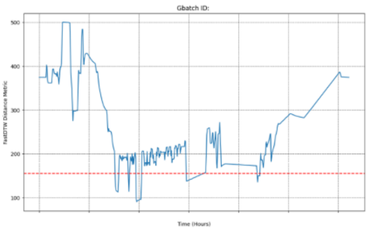
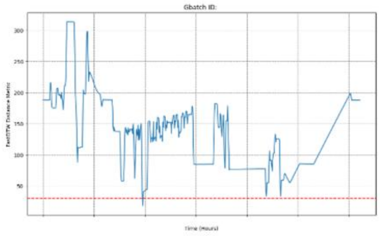
2nd Batch		
3rd Batch		
4th Batch		

Table 6 Demonstration of how removing the most variable feature facilitates the threshold setting

6. Testing

The testing phase consists of comparing the stored labelled results that are derived from the visual identification of all the manufacturing steps of the toothpaste process and the steps identified by the Data Science Model that has just been integrated with the new process derived from the mixer 2 at Oak Hill manufacture site.

This approach makes it possible to verify the accuracy of the Data Science Model in detecting steps by measuring what is the difference measured in minutes between manually identified steps and those automatically detected by the model allowing the identification of the required corrections and making the model more reliable. Generally, during the testing phase, changes were made to the Manufacturing Pattern Templates of the steps that showed higher discrepancy using the table shown in 2. Creation of Manufacturing Pattern Templates as a guide, or changes in the threshold setting. In addition, it is important to highlight that is decided a threshold value equal to the sum of all step differences in the process that determines whether or not the testing phase is passed.

During the testing phase, a problem was identified where the Fast DTW sometimes confused the start of a step due to a strong similarity of feature characteristics defining the same non-aqueous manufacturing process, as illustrated in the following image. Consequently, this step initially did not pass the testing phase. However, by adjusting the manufacturing template to make it more specific, and by a re-evaluation of a threshold the issue was fixed.

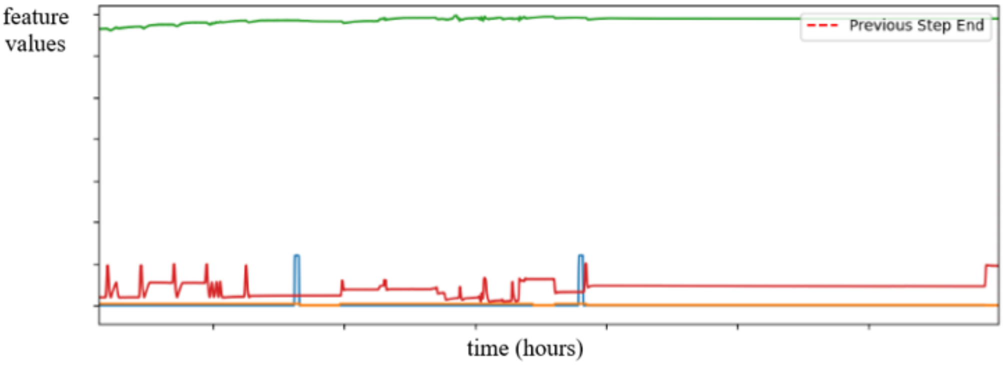


Fig. 43 Two similar phases in the non-aqueous manufacturing process

7. Implementation

Once the testing phase was passed, it was ensured that the model identifies most of the steps reliably, and so the next step is to implement online the model. During software implementation step, the software project needs to be inserted in the software system architecture, in particular, the piece of code needs to be integrated with the other modules ensuring that the model functionalities work correctly.

8 Conclusion: essential aspects of scalability management

To successfully tackle the development and scaling of a smart control system, it is crucial to adopt a balanced approach that gives equal priority to both the technical and management aspects.

From my personal experience in Haleon during the development of this project, from a managerial insight, it has proven to be of utmost importance not to neglect the needs of any key stakeholders (data scientists, process engineers, digital architects, and operators). For example, receiving feedbacks from operators is not a given because usually the working hours are different often due to the different time zone when the manufacture site is located in a different geographical area, but it is highly important to get their feedback so that the risks arising from daily activities can be minimized. This approach ensures that the solution is well integrated into the operational environment. Another important aspect is the strategic choice of the minimum viable product (MVP) which sets the foundation for successful scaling. The MVP choice is context-dependent, influenced mainly by the goals and the project complexity. Its importance derives mainly from the need to address the specific challenge of this project to develop and implement a digital solution simultaneously with the aim to accelerate the adoption and delivery value. In this specific case, the choice was made by exploiting the similarity of the production process of toothpaste building the version of the MVP on the basis of the most produced and with a higher profit margin toothpaste characterized by enough common aspects with the new processes and sites, thus facilitating the scaling process. It is also important to highlight that a well-defined goal makes the choice of the MVP more strategic making the path towards the success less challenging. From a technical standpoint, the challenge is greatly facilitated by a well-designed digital architecture characterized by high flexibility and able to manage a huge amount of data allowing an easier integration of new elements without compromising the overall performance. In addition, a scalable data science model that is based on an algorithm such as the Fast Dynamic Time Warping greatly enables it to support the integration of other processes by being made up of logic that is well suited to monitor any process regardless of their nature.

Bibliography

[1] "Haleon website," [Online]. Available: https://www.haleon.com/.

[2] "Haleon Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Haleon.

[3] "Organization Agility," [Online]. Available: https://scaledagileframework.com/organizational-agility/.

[4] "Proof of Concecpt, Prototype, Pilot, MVP what's in a name?," [Online]. Available: https://www.nesta.org.uk/blog/proof-of-concept-prototype-pilot-mvp-whats-in-a-name/.

[5] "Innovation practice from prototype to mvp," [Online]. Available: https://www.pentalog.com/blog/tech-trends/innovation-practice-from-prototype-to-mvp/.

[6] "Project Management Methodologies," [Online]. Available: https://www.forbes.com/uk/advisor/business/project-management-methodologies/.

[7] A. D. Marco, in *Project management for facility constructions* , 2018, p. 88.

[8] "Project Management Methodologies," [Online]. Available: https://asana.com/resources/project-management-methodologies.

[9] "Waterfall vs. Agile vs. Kanban vs. Scrum: What's the difference?," [Online]. Available: https://asana.com/resources/waterfall-agile-kanban-scrum.

[10] "Waterfall vs. Agile vs. Kanban vs. Scrum: What's the difference?," [Online]. Available: https://asana.com/resources/waterfall-agile-kanban-scrum.

[11] "What is a data Lake?," [Online]. Available: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake/#data-lake-definition.

[12] "Unleashing the power of Real-Time Machine Learning to accelerate the production of toothpaste in Haleon," [Online]. Available: https://medium.com/trusted-data-science-haleon/unleashing-the-power-of-real-time-machine-learning-to-accelerate-the-production-of-toothpaste-in-70ab7c3cf1f1.

[13] "FastDTW in Action: Optimising Manufacturing Operations," [Online]. Available: https://medium.com/trusted-data-science-haleon/fastdtw-in-action-optimizing-manufacturing-operations-c07f3cc5023c.

[14] “Dynamic time warping 1: Motivation,” [Online]. Available: https://www.youtube.com/watch?v=ERKDHZyZDwA.

[15] “Dynamic time warping 2: Algorithm,” [Online]. Available: https://www.youtube.com/watch?v=9GdbMc4CEhE.

[16] “Scaling Data Science,” [Online]. Available: https://www.oracle.com/it/a/ocom/docs/oracle-ds-scaling-data-science.pdf.

[17] “Come Usare Git e GitHub,” [Online]. Available: https://www.programmareinpython.it/blog/come-usare-git-e-github/.