Artificial Intelligence for Portfolio Management Optimization - "Portfolio Squeeze"

Matheus Canciglieri
Leonardo Piovesan
Marcelo Viegas
Robert Bosch Ltda

Abstract

In the era of Industry 4.0, businesses are undergoing a radical transformation, reshaping operations through the convergence of cyber-physical and digital systems. Central to this transformation is the pivotal role of artificial intelligence (AI), specifically the application of clustering techniques, in enhancing portfolio management. The complexity of modern products and services demands strategic resource allocation, minimizing risks and maximizing returns. This study explores the potential of AI, particularly Agglomerative Clustering, to optimize the analysis of "S" nozzle portfolio. Portfolio management, crucial for success in the current business landscape, faces challenges in making precise decisions amid vast information and variables. The study applies a qualitative, experimental case study approach, demonstrating the robustness of the methodology. The chosen Agglomerative Clustering algorithm's hierarchical and incremental clustering proves effective in handling complex data. The literature review emphasizes the growing importance of portfolio management, while clustering, particularly Agglomerative Clustering, has proven efficacy in diverse areas. The results reveal that 76 out of 430 nozzles are eligible for LT50 at CtP plant, with an overall eligibility percentage of 17.7%. Different nozzle types exhibit varying eligibility rates, emphasizing the effectiveness of the methodology. In conclusion, the integration of AI, specifically clustering techniques, offers a substantial contribution to portfolio management in the industry 4.0 era. This study provides a significant step in practically applying these technologies, offering insights for future research and implementations in the dynamic context of portfolio management.

1. Introduction

The era of Industry 4.0 has been radically reshaping how companies operate and engage with technology. With the convergence of cybernetic, physical, and digital systems, we witness a revolution in production, management, and decision-making. In this scenario, artificial intelligence (AI) emerges as a fundamental catalyst, empowering organizations to optimize processes, enhance efficiency, and explore unprecedented opportunities. This technological transformation not only redefines productivity standards but also drives the evolution of sectors such as portfolio management, where the strategic application of AI is increasingly crucial.

Portfolio management plays a paramount role in today's business context. With the increasing complexity and variety of products and services offered by an organization, the ability to efficiently manage a portfolio of projects, investments, or products becomes decisive for market success as part of the product lifecycle management (PLCM). The capacity to strategically select and prioritize available resources to maximize returns while minimizing associated risks is an essential competitive advantage for modern companies.

However, portfolio management faces significant challenges. Making precise and swift decisions becomes complex in the face of the vast amount of information and variables involved. Identifying investment opportunities, allocating resources effectively, and managing risks are tasks that demand not only human expertise but also computational power capable of intelligently processing large volumes of data.

In this context, the use of artificial intelligence, particularly the application of clustering techniques, emerges as a promising solution to enhance portfolio management. AI's ability to identify patterns, group similar data, and extract valuable insights from complex information sets offers a significant strategic advantage. Clustering allows the segmentation of portfolio data into homogeneous groups, facilitating analysis and enabling more precise and informed decision-making.

The purpose of this article is to explore the possibilities of artificial intelligence, specifically the application of clustering techniques, to optimize portfolio management and leverage the advantages of scalability. The goal is to provide a thorough analysis of how AI can be effectively and strategically integrated into portfolio management, with the aim of improving resource allocation, increasing profits, and mitigating risks associated with "S" type portfolios.

2. Literature Review

2.1. Custerization

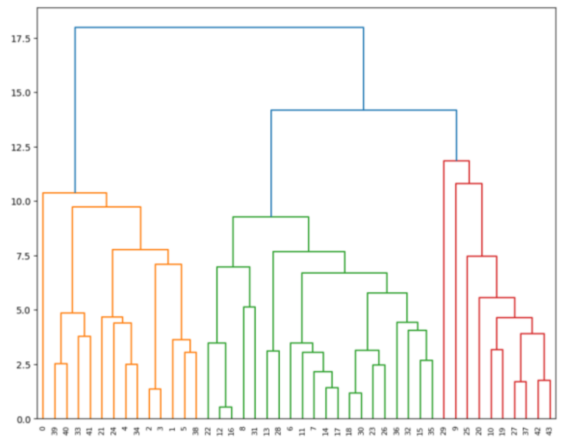
Clustering is an essential data analysis technique aimed at grouping similar elements into distinct sets, called clusters, based on their common characteristics. This approach has wide application in various fields such as pattern recognition, image processing, bioinformatics, among others.

Among the clustering methods, Agglomerative Clustering is a hierarchical approach often used. This method starts by considering each data point as an isolated cluster and then iteratively
combines the closest clusters to each other based on some distance metric, such as Euclidean distance. This process continues until all data points are grouped into a single cluster or until a predefined number of clusters is reached.

Agglomerative Clustering methods operate iteratively, merging the most similar clusters at each step of the process. There are different methods for calculating the distance between clusters, such as Single Linkage, Complete Linkage, and Average Linkage. For example, Single Linkage considers the shortest distance between the closest points from different clusters, while Complete Linkage uses the longest distance between the farthest points from distinct clusters.

One advantage of Agglomerative Clustering is the ability to construct dendrograms (Figure 2), visual representations that show the hierarchy of data grouping. This can be useful for determining the ideal number of clusters by observing where a significant jump in distance (height) occurs in the dendrogram.

Figure 1. Exemple of Dendogram.



In a recent study by Zhang et al. (2021), Agglomerative Clustering was successfully applied to customer segmentation for analyzing purchasing behavior, allowing for a deeper understanding of consumption patterns. Another study by Wu et al. (2019) explored the application of Agglomerative Clustering in gene expression analysis, identifying gene expression profiles under different physiological conditions.

However, despite its effectiveness in identifying hierarchical structures, Agglomerative Clustering may face challenges in large datasets due to its computational complexity. To overcome this, recent research, such as the work of Charrad et al. (2020), has proposed optimization strategies and performance enhancement of Agglomerative Clustering on large-scale datasets.

In summary, Agglomerative Clustering is a valuable and widely used method in data analysis for grouping information hierarchically. Its use has extended to various areas, demonstrating potential in different application contexts, although scalability challenges remain the subject of research in recent studies.

2.2. Portfolio Management

Portfolio management is an essential strategic process for organizations seeking to optimize their investments, resources, and projects. This practice involves the selection, prioritization, and control of a set of projects or investments aligned with the company's strategic objectives. In recent years, the literature on portfolio management has focused on various approaches and methodologies to improve the efficiency and performance of this process. Authors such as Cooper et al. (2019) explore the importance of portfolio management as a tool for aligning investments with the organization's strategy, highlighting the need for prioritization based on strategic criteria. Furthermore, recent studies by Lechler et al. (2020) emphasize the integration of portfolio management with resource management, highlighting the importance of allocating resources efficiently to maximize the value added to projects. The ability to optimize resource allocation according to portfolio demands is a central point in recent literature on the subject. Another relevant aspect discussed in recent literature is the application of agile methodologies in portfolio management. Authors such as Lahrmann et al. (2018) explore how the implementation of agile practices can increase portfolio flexibility and responsiveness to changes, improving adaptation to dynamic environments. Technology also plays a significant role in modern portfolio management. Studies such as Archibald et al. (2021) address the use of digital tools and platforms to facilitate portfolio management and analysis, enabling more informed and agile decision-making.

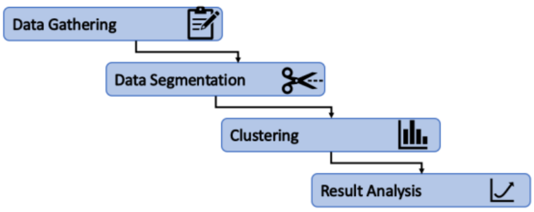
Furthermore, the integration of data-driven methodologies and predictive analysis has been a growing focus in portfolio management. Authors such as Shenhar and Dvir (2020) highlight the importance of advanced analytical techniques for predicting risks and opportunities, assisting in making more informed decisions and proactively identifying challenges that may arise in the portfolio.

In summary, portfolio management has evolved considerably in recent years, driven by more strategic approaches, integration of agile methodologies, adoption of digital technologies, and the application of predictive analytics. These trends reflect the growing need for organizations to maximize the value of their investments and resources, ensuring that the portfolio is aligned with strategic objectives and capable of adapting to market and business environment changes.

3. Methodology

The research project developed is of an applied nature since it utilizes concepts and methodologies already developed and validated to create practical knowledge with the aim of achieving the proposed objectives. Its approach is considered qualitative as it seeks a deep understanding of a specific phenomenon, analyzing variables that cannot be numerically measured (such as information consistency and conceptual relationships). Finally, the technical procedures follow the model of an experimental case study (Figure 2).

Figure 2. Technical research procedures.



3.1. Data Clustering

For data manipulation, the Pandas and Numpy libraries were used, known for their robust capabilities in information processing. A total of 430 distinct Part Numbers (PN) of S nozzles were identified, where the following information was mapped: i) Injector hole diameter; ii) Injector hole length; iii) Minimum Hydraulic Flow; iv) Maximum Hydraulic Flow; v) Needle Stroke; vi) Snoring group; vii) Blind hole diameter; viii) Blind hole length; ix) Seat measurement; x) Dimension A; xi) Pin diameter; xii) Pin length; xiii) Rod length; xiv) Total rod length; xv) Phi angle; xvi) Psi angle; xvii) Type; xviii) Mounting figure; and xix) Number of holes. Before implementing the clustering algorithm, a preliminary segmentation of the dataset was performed. This preliminary step was based on criteria discussed and outlined in meetings with specialists in the specific field. The criteria considered for this preliminary division encompassed attributes that could not be directly consolidated, such as the number of holes, as well as the type and configuration of fixing. Table 1 presents the quantity of PN referring to each subdivision.

Table 1. Pre-Division of the PNs.

Hole	Type	Fixation Figure	PN
1	DLLA	5	1
3	DLLA	2	1
3	DLLA	5	19
4	DLL	1	3
4	DLL	2	10
4	DLL	5	3
4	DLL	7	3
4	DLL	8	1
4	DLLA	5	206
4	DLLA	8	1
4	DLLB	6	5
4	DLLZ	14	2
5	DLL	2	5
5	DLLA	5	80
5	DLLB	6	1
5	DLLC	6	1
5	DLLZ	14	4
6	DLL	2	3
6	DLLA	5	28
7	DLL	2	2
7	DLL	5	1
7	DLL	16	3
7	DLLA	5	10
7	DLLZ	14	4
8	DLL	2	6
8	DLL	5	2
8	DLLA	5	16
8	DLLZ	14	3
9	DLLZ	14	2
12	DLLA	5	2
12	DLLA	8	1
12	DLLZ	14	1

The initial segmentation phase was crucial for establishing categories or initial groups in the data, with the purpose of providing relevant insights during the subsequent clustering phase. The detailed analysis of these specific attributes allowed the identification of primary characteristics that could influence the formation of the final clusters, contributing to a more refined and meaningful approach in the subsequent analysis. Due to the nature of clustering algorithms, divisions that only contained one PN were excluded from the process. Each part number (PN) identifier had its phi and psi angles sorted in ascending order, an approach that resulted in significant improvements in test results without causing adverse practical implications. The variables selected for the clustering process were determined based on their relevance, and the following were chosen: Minimum Hydraulic Flow, Maximum Hydraulic Flow, Phi and Psi Angles, Rod Length, and Total Rod Length.

To ensure uniformity in the scale of the data, the StandardScaler function from the scikit-learn library was employed for data standardization. This procedure resulted in transforming the data into a distribution with zero mean and unit standard deviation, enabling fair comparison among the considered magnitudes.

The selected algorithm to conduct the clustering process was Agglomerative Clustering, available in the scikit-learn library. This algorithm was chosen due to its hierarchical functioning principle, which allows starting with all data grouped into a single cluster and gradually separating them until each data point belongs to an exclusive cluster. A notable feature of this algorithm is its ability to incrementally build the grouping structure, which is especially advantageous in situations where hierarchy in groupings is a relevant factor.

For these groups, the distance criteria known as 'ward' was used. This criteria is employed to minimize variance between the clusters being grouped. Essentially, 'ward' seeks to merge clusters in a way that minimizes the sum of squares of differences between all points in the clusters, aiming to create more cohesive and compact groupings. This approach is valuable for identifying patterns and structures in the data, contributing to a more precise and meaningful interpretation of the resulting groupings.

To conduct the analysis, the outcomes of the models were organized into tables of variations, providing information about the maximum variation present for each variable within each of the identified clusters. These tables allowed for a thorough evaluation of the discrepancies observed in each cluster, highlighting the maximum variations recorded for each considered variable.

Subsequently, an important step was the counting of the number of clusters from each model that fully satisfied the established variation criteria. These criteria were defined as: variation in flow (Δ flows) less than or equal to 10%, variation in rod length (Δ Rod Length) less than or equal to 0.125 mm, and variation in angles (Δ Angles) within the limit of 5°.

This approach allowed for a precise and quantitative assessment of the consistency of the clusters formed by each model concerning the stipulated variation criteria. Counting the number of clusters that fully met these criteria provided a clear view of the robustness and adherence of the models, offering crucial information for understanding the behavior and validity of the generated clusters.

Finally, tables were generated for each model's result (Table 2), detailing the clusters that met the variation criteria and the count of PNs in each cluster. The calculation to determine the eligible PNs was performed by subtracting the sum of the quantity of PNs in the clusters by the quantity of clusters that satisfied the variation criteria. This metric is crucial as it seeks to maintain only one active PN per cluster, deactivating the others, simplifying representation, and reducing redundancies.

The model selected for data clustering was determined based on the quantity of eligible PNs. The one that presented the highest number of eligible PNs was chosen, indicating its capacity to offer a more concise and effective representation of the data, in accordance with the stipulated criteria. This criterion was fundamental for selecting the most suitable model for data clustering.

4. Results

The outputs of the hierarchical clustering approach not only provided crucial data but also played a fundamental role in formulating an analytical method, the primary purpose of which is to significantly optimize the clustering process. This method was conceived based on the relentless pursuit of the ideal cluster configuration, prioritizing the maximization of the number of eligible Parts Numbers (PNs) for LT50, while strictly adhering to the established variation criteria.

The adopted strategy focused meticulously on identifying the cluster configuration that not only offered the highest possible number of PNs eligible for LT50 but also enhanced the representation of each cluster. This allowed for a more precise and meaningful identification of PNs that could be viable for LT50, thus contributing to a more detailed and selective understanding of the analyzed process.

Upon observing Table 2, it is evident that out of 430 nozzles, 76 were identified as eligible for LT50, representing an overall eligibility rate of 17.7%. When segmenting the analysis by hole type, a substantial variation is observed, ranging from 10.0% to 50.0%. The "DLLZ" type with mounting figure 14 stands out with the highest eligibility percentage, reaching 33.3%. There is also a variety in the eligibility rates of the "DLLA" type with mounting figure 5, with the situation with 4 holes being particularly notable, presenting 50 eligible PNs, equivalent to 24.3% of the total in this group.

After the initial phase of creating the clusters, the analysis of similarity among elements in each grouping played a crucial role. This stage aimed to identify the most promising clusters for subsequent testing and validations. The similarity analysis provided valuable insights into existing relationships and patterns within each cluster, being essential for optimizing the subsequent testing phases, focusing on sets of elements that share more prominent intrinsic relationships.

The comprehensive methodology employed in the cluster analysis, which includes the assessment of true clusters, the calculation of LT50 eligibles, manual analysis of false clusters, and similarity evaluation, significantly contributes to a precise and selective understanding of the analyzed process.

Table 2 – Portfolio Optimization Results

Holes	Type	Fixation Figure	PN	Eligible PN	Eligible %
4	DLLA	5	206	50	24,3%
5	DLLA	5	80	10	12,5%
6	DLLA	5	28	5	17,9%
3	DLLA	5	19	3	15,8%
8	DLLA	5	16	3	18,8%
7	DLLA	5	10	2	20,0%
4	DLL	2	10	1	10,0%
8	DLLZ	14	3	1	33,3%
12	DLLA	5	2	1	50,0%
TOTAL			430	76	17,7%

The final step involved analyzing the similarities between the elements of each cluster. This analysis was conducted on normalized variables, ensuring that the lowest value in each cluster was zero. By normalizing the variables, we effectively standardized the scale, facilitating a fair comparison across clusters. The similarity analysis then compared the highest and lowest value elements in each cluster. The closer to zero this difference was, the more similar the elements were deemed to be, indicating a higher likelihood of yielding favorable outcomes in practical applications. The resulting similarities are detailed in Table 3, providing a comprehensive overview of the clustering outcomes, and highlighting the degrees of similarity within and between clusters. As can be seen in Table 3, the best cluster has a similarity of 99,8% while the worst cluster still has a similarity of approximately 60%. This cohesive integration between advanced analytical methods and practical results solidifies the adopted approach, providing valuable insights and informing informed decisions for continuous process optimization.

Table 3. Similarity Analysis

Cluster	Part Numbers					Similaridade
12F DLLA F5 C0	433272980	433272985				99,80%
4F DLLA FF5 C24	433271589	433271471				99,22%
5F DLLA F5 C14	433271521	433271587				98,25%
8F DLLA F5 C5	433271635	433271632				97,89%
5F DLLA F5 C12	433271695	433271674				97,75%
6F DLLA F5 C4	433272961	433272963	433272960			96,82%
4F DLLA FF5 C16	433271418	433271430				96,28%
4F DLLA FF5 C5	433271773	433271725				96,26%
5F DLLA F5 C13	433271682	433271658				95,97%
4F DLLA FF5 C19	433271212	433271240				95,81%
4F DLLA FF5 C20	433271180	433271195				95,34%
8F DLLA F5 C2	433271459	433272004				95,30%
4F DLLA FF5 C47	433271591	433271487				94,96%
4F DLLA FF5 C6	433271755	433271734				94,73%
5F DLLA F5 C6	433271571	433271588	433271590			94,16%
4F DLLA FF5 C30	433271838	433271882				93,27%
5F DLLA F5 C5	433271756	433271713				93,27%
4F DLLA FF5 C18	433271759	433271763				92,61%
8F DLLZ F14 C0	433271637	433271651				92,55%
4F DLL FF2 C0	433220140	433220109				92,53%
4F DLLA FF5 C28	433271433	433271434				92,39%
4F DLLA FF5 C34	433271395	433271301				92,08%
6F DLLA F5 C7	433271525	433271526				91,46%
5F DLLA F5 C0	433271837	433271868				91,28%
5F DLLA F5 C3	433271729	433271692	433271753			90,87%
4F DLLA FF5 C35	433271223	433271377				90,29%
4F DLLA FF5 C72	433271842	433271780				90,03%
4F DLLA FF5 C36	433271263	433271736				89,61%
4F DLLA FF5 C3	433271819	433271711	433271754			89,13%
6F DLLA F5 C2	433271656	433271655				88,92%
4F DLLA FF5 C21	433271774	433272968	433271796			88,69%
7F DLLA F5 C1	433271603	433271604				88,30%
4F DLLA FF5 C22	433271772	433271818				87,77%
4F DLLA FF5 C25	433271486	433271485	433272012			87,15%
3F DLLA FF5 C2	433271775	433271768	433271643	433271795		86,21%
4F DLLA FF5 C45	433271512	433271789				86,00%
7F DLLA F5 C4	433271647	433271678				84,87%
4F DLLA FF5 C55	433271823	433271836				84,29%
5F DLLA F5 C29	433271345	433271373				82,81%
4F DLLA FF5 C13	433271811	433271856				80,96%
4F DLLA FF5 C74	433271865	433271280				80,28%
4F DLLA FF5 C40	433271260	433271309	433271425			80,15%
4F DLLA FF5 C29	433271501	433271722				79,53%
4F DLLA FF5 C46	433271503	433271499	433271495			79,37%
4F DLLA FF5 C54	433271834	433271879				79,09%
4F DLLA FF5 C23	433271357	433271686				78,47%
4F DLLA FF5 C10	433271887	433271884				76,32%
4F DLLA FF5 C15	433271891	433271849	433271824			76,24%
4F DLLA FF5 C9	433271352	433271205				74,26%
4F DLLA FF5 C27	433271313	433271423				73,60%
4F DLLA FF5 C11	433271864	433271806				73,50%
4F DLLA FF5 C7	433271848	433271871	433271417	433271431	433271043	73,46%
4F DLLA FF5 C12	433271214	433271451	433271494			72,27%
6F DLLA F5 C6	433271602	433271668				69,03%
4F DLLA FF5 C1	433271366	433271489	433271800			67,48%
8F DLLA F5 C0	433271672	433272044				65,42%
4F DLLA FF5 C4	433271360	433271816				64,23%
4F DLLA FF5 C26	433271101	433271217	433271411			59,81%
4F DLLA FF5 C0	433271348	433271709				59,52%

5. Conclusion

As Industry 4.0 profoundly reshapes business practices, portfolio management emerges as a critical area for organizational success. The increasing complexity of products and services demands a strategic approach to resource allocation, minimizing risks, and maximizing returns. In this context, artificial intelligence (AI), particularly clustering techniques, presents itself as a promising solution to enhance portfolio management.

This study explored the potential of AI, using Agglomerative Clustering, to analyze and optimize a specific dataset related to "S" nozzles. The applied methodology demonstrated a robust approach, involving careful pre-division of the data and effective implementation of the clustering algorithm. The choice of Agglomerative Clustering method, based on its hierarchical and incremental capability, proved suitable for handling data complexity.

Literature review highlighted the growing importance of portfolio management, while clustering, especially Agglomerative Clustering, proved effective in various areas, including customer segmentation and gene expression analysis.

The results obtained through cluster analysis provided valuable insights for identifying promising groupings. The strategy of counting the number of clusters that met the established criteria allowed for a quantitative assessment of the models' robustness. The final analysis revealed that 76 out of 430 nozzles were eligible for LT50, with an overall eligibility percentage of 17.7%. The situation of "DLLA" nozzles with fixing figure 5, where 4 holes showed a notable eligibility rate of 24.3%, highlights the method's effectiveness.

The AI's ability to identify patterns, group data, and extract insights was crucial to enhance portfolio management. By segmenting data into homogeneous clusters, the approach allowed for a more precise and informed analysis, guiding strategic decision-making.

In summary, the integration of artificial intelligence, especially through clustering techniques, offers a substantial contribution to portfolio management, providing an advanced and effective analytical approach. This study represents a significant step in the practical application of these technologies, offering a foundation for future research and implementations in the dynamic context of portfolio management in the Industry 4.0 era.

6. References

Archibald, R. D., Kerzner, H., & Kharbanda, O. P. (2021). Managing high-technology programs and projects (4th ed.). Wiley.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2020). NbClust: An R package for determining the relevant number of clusters in a data set. Journal of Statistical Software, 61(6), 1–36. https://doi.org/10.18637/jss.v061.i06

Cooper, R. G., Edgett, S. J., & Kleinschmidt, E. J. (2019). Portfolio management for new products (3rd ed.). Basic Books.

Lahrmann, G., Marx, F., Winter, R., & Wortmann, F. (2018). Business intelligence maturity: Development and evaluation of a theoretical model. Information Systems Management, 28(4), 239–254. https://doi.org/10.1080/10580530.2011.607181

Lechler, T., Edington, B., & Gao, T. (2020). Project management in the strategy execution process. Journal of Management in Engineering, 28(4), 1–11. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000104

Shenhar, A. J., & Dvir, D. (2020). Reinventing project management: The diamond approach to successful growth and innovation. Harvard Business Review Press.

Wu, H., Li, X., Xu, L., & Zhang, Y. (2019). Gene expression data analysis using agglomerative hierarchical clustering with adjusted distance metric. BMC Bioinformatics, 20(1), 1–12. https://doi.org/10.1186/s12859-019-2998-4

Zhang, J., Wang, L., & Chen, Y. (2021). Customer segmentation using agglomerative clustering for targeted marketing. Expert Systems with Applications, 168, 114–124. https://doi.org/10.1016/j.eswa.2020.114124