Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v10. n.50, p. 156-167. November/ December., 2024.

DOI: https://doi.org/10.5935/jetia.v10i50.1262

RESEARCH ARTICLE OPEN ACCESS

USING MACHINE LEARNING TO EVALUATE INDUSTRY 4.0 MATURITY: A COMPREHENSIVE ANALYSIS HIGHLIGHTING LEAN'S IMPACT ON DIGITAL TRANSFORMATION

Oussama Ben Ali[1], Sondes Hammami[2], Marwa Hasni[3], Fehmi H'Mida[4], Ahmed Nait Sidi Moh[5]

[1]National Engineering School of Tunis, University of Tunis El Manar, Tunisia
[2]National Engineering School of Tunis
[3]National Engineering School of Tunis, University of Tunis El Manar, Tunisia
[5] National Engineering School of Carthage, Université of Carthage, Tunisia
[4] Jean Monnet University, LASPI Laboratory, Roanne, France.

[1] http://orcid.org/0009-0000-6568-5387, [2] http://orcid.org/0000-0002-0044-1338, [3] http://orcid.org/0000-0002-4509-5563
[4] http://orcid.org/0009-0001-7009-823X, [5] http://orcid.org/0000-0002-2297-8603

Email: benalioussama1997@gmail.com, sondes.hammami@enicar.ucar.tn, marwa.gharbi@enib.ucar.tn, mohamedfehmi.hmida@enicar.ucar.tn, ahmed.nait@univ-st-etienne.fr

ARTICLE INFO

Article History
Received: August 21, 2024
Revised: October 1, 2024
Accepted: November 10, 2024
Published: December 31, 2024

Keywords:
Digital Transformation · Lean · Machine Learning · Industry 4.0 maturity. Data Analysis · Maturity level prediction

ABSTRACT

The rise of digital technologies in manufacturing and industries, known as the fourth industrial revolution, has created both opportunities and challenges for businesses. To succeed in this era of "Industry 4.0," companies need to assess their digital maturity. Through this study, we analyze the global state of Industry 4.0 maturity, identifying industry-specific trends, challenges, and potential growth. Leveraging advanced machine learning techniques, including data analysis, prediction, and recommendations. The study explores the complexities and evolution of Industry 4.0. Additionally, we show how machine learning plays a pivotal role in this analysis, contributing to enhanced insights and decision-making capabilities. Our research aims to not only assess the current state but also forecast future roadmaps while providing tailored recommendations for enhancing maturity levels. We aim to evaluate various machine learning based approaches for addressing these inquiries, focusing on Decision Tree, Support Vector Machine, and Random Forest models. We will choose the best performing model for our scenario. Initially, we use unsupervised learning through Hierarchical Clustering for grouping data, followed by data expansion. Subsequently, we employ supervised learning techniques, particularly Decision Tree, for descriptive, predictive, and perspective analysis. Among our recommendations for enhancing Industry 4.0 maturity levels, we advocate for extensive interventions, but exclusively for companies meeting predetermined criteria delineated within the decision tree node. Furthermore, we examine the influence of Lean on digital transformation. Through this interdisciplinary approach, our findings contribute to a deeper understanding of Industry 4.0 evolution and offer practical insights for strategic decision-making in the era of digitalization.

I. INTRODUCTION

The growth of Industry 4.0, often linked with the idea of the "smart factory" involves using high-tech tools like the Internet of Things (IoT), artificial intelligence (AI), advanced robots, 3D printing, and cloud computing in business operations. This shift completely changes how companies work and connect with their surroundings. It represents a new era in manufacturing marked by flexibility, creativity, and sustainability.

According to [1], there are five main reasons why Industry 4.0 is so important and groundbreaking: (1) Businesses can adapt quickly to market changes thanks to automation, (2) It boosts innovation and productivity, (3) It puts consumers at production, demanding new skills, and (5) it is posited to foster sustainable prosperity by leveraging modern technologies to address energy, resource, environmental, and socio-economic challenges.

Therefore, companies aim to align with the emerging trend of Industry 4.0 to maintain competitiveness in the market. To achieve this, it is essential to assess current levels of maturity and develop a clear roadmap for improvement.

Thus, companies stand to gain valuable insights by determining factors for their maturity level within the context of Industry 4.0. More particularly, we assess whether a company with a high maturity level in Lean is better positioned to enhance its Industry 4.0 maturity for becoming an advantageous endeavor. In other words, the process of improving the maturity level of Industry 4.0 could be smoother (quicker and less costly for the company) if it already has an established maturity basic level in this area.

To the best of our knowledge, there hasn't been a study that specifically outlines the key factors affecting the maturity level of Industry 4.0. While many studies have examined the correlation between Lean production and Industry 4.0 [2], [3], [4] they often neglect to specify how the maturity level of Lean impacts that of Industry 4.0.

The authors of [4] interviewed several companies to investigate the relationship between Lean practices and Industry 4.0. This investigation covered three main areas: Lean practices, Industry 4.0 technologies, and company performance indicators. The study revealed that combining Lean practices with Industry 4.0 adoption leads to significant improvements in operational performance. Additionally, it was observed that Lean practices were prevalent in companies with high operational performance improvement, while the adoption of Industry 4.0 was not significantly associated with such improvements. Despite this, there was a notable linkage between Industry 4.0 adoption and Lean practice implementation. While this study emphasizes the relationship between Lean Production, Industry 4.0 and operational performance, it does not show a tailored roadmap and improvement recommendations to companies. This gap is, indeed, undertaken in our current paper.

An additional study in [3] designed a questionnaire to categorize companies based on their commitment to Lean practices. The study discovered that companies that are deeply committed to Lean practices approached digital transformation differently from those with lower commitments. As such, two distinct digital transformation patterns have been identified: the Sustaining pattern, characterized by gradual digitalization involving the entire company horizontally. And the Disruptive pattern, marked by significant digital investments with a vertical focus. This research underscores the importance of understanding different digital transformation strategies for practitioners and scholars. Although this article indicates the difference between companies with low Lean level and high Lean level in adopting industry 4.0 technologies.

The contribution of other factors and their impact on industry 4.0 maturity level is not discussed.

Our proposed model allows to identify and analyze these factors. As a result, our approach is characterized by its customized and dynamic nature, that is meticulously aligned with the unique objectives and needs of the company.

Going into more details, Artificial Intelligence (AI) is used to identify the critical factors influencing the maturity level of Industry 4.0, particularly by assessing the impact of Lean on it. We have opted for AI due to its ability to explore complex patterns, detect nonlinear relationships, and adapt to evolving data [5]. Moreover, AI enables us to develop predictive and prescriptive models to anticipate future trends and recommend strategic actions.

In summary, our AI-based approach offers a powerful and adaptable solution for understanding and enhancing the maturity level of Industry 4.0 in businesses.

The aim of our study is to conduct an analysis of the primary factors influencing the enterprise maturity level on Industry 4.0. Additionally, we are seeking to ascertain whether the maturity level of an enterprise in Lean practices affects its maturity level in Industry 4.0. We will also explore the existence of maturity models specifically tailored for Digital Lean initiatives.

Furthermore, we intend to identify suitable machine learning methodologies to address these inquiries. First, we use an unsupervised ML technique for data clustering and data augmentation which is Hierarchal clustering. Then, we use supervised ML techniques for the analysis issue. We mainly explore Decision Tree, Support Vector Machine and Random Forest for predictive analysis and Decision Tree for descriptive and perspective analysis.

The rest of this paper is organized as follows. In Section 2, we give a background and literature review regarding the addressed subject. Section 3 gives Dataset and methodology and in section 4, we represent results and discussion.

II. BACKGROUND

Before delving into our research questions, it is pertinent to establish a comprehensive definition of a maturity model. Subsequently, we will provide an overview of prominent maturity models relevant to Industry 4.0, Lean methodologies, and ascertain the existence of maturity models focused on digital Lean transformations. To this end, we will discuss potential examples of machine learning tools that could be explored to analyze and identify the key factors influencing the maturity level of enterprises in Industry 4.0 initiatives.

II.1 MATURITY MODEL

According to [6], maturity models are commonly used to assess the current situation, identify and prioritize improvement measures, and monitor progress in a given domain. These models are designed as a set of levels or stages describing the development of the examined object in a simplified manner, as described by [7].

Besides, the study in [8] highlights that maturity models allow us to define the current and desired levels of maturity along with corresponding improvement measures. In the same way, [9] describe maturity models as tools for continuous improvement and guides for organizations.

While maturity models may vary in their structural design and application domains, they typically comprise two fundamental components to serve their intended purpose as given in [6], [8]: (1) a series of levels or stages, and (2) dimensions or capabilities.

Dimensions serve as pivotal points in the evaluation process, with their selection tailored to the specific domain being assessed. Maturity models are inherently multidimensional. For example, in assessing the maturity of an enterprise in the context of Industry 4.0, considerations extend across various dimensions such as processes, personnel, and technology [10].

Similarly, when evaluating enterprises based on the maturity of their supply chain, dimensions such as reverse logistics, collaboration, processes, technology, and sustainability come into play [11]. Thus, the choice of dimensions to be evaluated is contingent upon the specific domain under scrutiny.

The authors of [6], [12], [8] classify maturity models into three categories based on their intended use:

- 1) descriptive, where the model is used to assess the maturity level of each dimension and the overall maturity level. Thus serving as a diagnostic tool to help companies focus on dimensions with mediocre maturity levels.

- 2) prescriptive, where the maturity model provides guidelines in the form of a roadmap to help the company improve its maturity level.

- 3) comparative, where the model allows for internal or external benchmarking with sufficient historical data from many assessment participants.

We note that maturity models are often associated within readiness models. They behave like maturity models but focus on assessing how prepared systems are for change. These models start by understanding the current state of the system, which helps in getting ready for improvements. They evaluate where a system stands before it undergoes any transformation towards maturity [13].

II.2 MATURITY MODEL IN INDUSTRY 4.0

In the context of Industry 4.0, maturity models play a crucial role. They help spread awareness of the concept and offer companies a better grasp of it. Additionally, they provide practical suggestions for implementing strategies to adapt to this transformative revolution. In this section, we focus on Industry 4.0 maturity and some readiness models developed in the literature.

Based on [13], maturity models are classified according two criteria:

1) the nature of the model's maturity (SIMMI 4.0 model [14]) or readiness (DREAMY model [15], FORRESTER model [12]).

2) the objective of the models: descriptive SIMMI 4.0 [14] and Impuls model [16], perspective models (the Connected Enterprise Maturity Model [17]), descriptive and perspective models (DREAMY model [15]), and finally comparative and perspective models.

For each model, the authors present dimensions. It's worth noting that dimensions vary significantly from one model to another. For instance, the SIMMI model considers Vertical integration, Horizontal integration, Digital product development, and Cross-sectional technology Criteria. On the other hand, The Connected Enterprise Maturity Model proposes dimensions such as Information infrastructure (hardware and software), controls and devices (sensors, actuators, etc., that feed and receive data), networks (that move all this information), and security policies (understanding, organization, enforcement).

Most models consist of five stages but with different labels. For example, SIMMI proposes these levels: Basic digitization level, Cross-departmental digitization, Horizontal and vertical digitization, Full digitization, and Optimized full digitization. While the Connected Enterprise Maturity Model proposes the levels: Basic digitization level, Cross-departmental digitization, Horizontal and vertical digitization, Full digitization, and Optimized full digitization.

Different methods are used to collect data for evaluating the maturity level, including an online self-assessment tool, a questionnaire combined with visits (Industrie 4.0 Maturity Index from ACATECH), and a general questionnaire, assisted by a third party (SIMMI 4.0).

According to [10], Industry 4.0 maturity models from the literature cannot adapt to small and medium-sized enterprises. They proposed an Industry 4.0 maturity model specifically tailored for small and medium-sized enterprises.

This model Comprises six dimensions and six levels. Based on the sub-dimensions, they developed a questionnaire and then they built a matrix to help them identify the gaps.

II.3. MATURITY MODEL IN LEAN

Lean methodology is based on the principle of continuous improvement. Maturity models play a crucial role in this context. By evaluating the current maturity level of Lean, companies can identify the gaps which could be the targeted areas for improvement. Accordingly, the maturity level of Lean can undergo continuous improvement.

Referring to [18], the discussion encompasses 24 key models. Within this collection, several models stand out as Lean Construction Maturity Models (LCMMs), explicitly illustrating the concept of Lean Construction (LC) maturity and offering a structured approach to assess LC maturity. Additionally, there is a subset of models that clearly consider Lean Construction as a foundational element. Lastly, three models were identified that lack a direct link to LC maturity but incorporate Lean principles and their adaptability into their frameworks [19].

Besides discuss combining Building Information Modelling (BIM), which is a process for creating and managing information on a construction project, and Lean Management (LM) to make construction projects more efficient. They agree that using BIM and Lean together can reduce waste and improve project results [20].

In addition, [21] presented LEAST which is a Lean Enterprise Self-Assessment Tool developed by an industry/government/academia team under the auspices of the Lean Aerospace Initiative (LAI). LEAST is organized into three sections:

- Section 1: lean transformation/leadership,

- Section 2: life cycle processes,

- Section 3: enabling infrastructure.

Each section is composed of multiple sub-sections and each sub-section contains multiple Lean practices. In total LEAST presents 54 Lean practices. For each practice, five maturity levels were assigned from level 1 (least capable) to level 5 (world class). Evaluating all the 2 will give the company a clear vision of its current state and the best way to prioritize the targeted areas for improvement measures.

In their research, [22] introduced a Lean maturity model tailored for operational-level planning. The study emphasizes the criticality of organizations implementing a comprehensive enterprise-wide Lean transformation plan, such as LESAT-LAI (Lean Enterprise Self-Assessment Tool) developed by the Lean Advancement Initiative (LAI), as a foundational requirement for the proposed model's effectiveness.

The model created has the dual capability of assessing both the degree of leanness and the effectiveness of Lean practices by evaluating Lean performance. The main purpose of this study is to develop a Lean maturity model adapted to the specifications of Manufacturing Cells.

Data is gathered through a case study approach involving two Manufacturing Cells, allowing for an investigation of Lean maturity in a real-life context. Both quantitative and qualitative data are analyzed inductively to enhance the theoretical framework, interpret Leanness and performance results, and formulate overall measurements.

II.4. MATURITY MODEL IN DIGITAL LEAN

Combining both paradigms will require a new maturity model that takes into account the relationship between Lean and Industry 4.0. In the relevant literature, only few articles have presented the Lean 4.0 maturity model. While there are articles proposing a Lean-Based Maturity Framework integrating Building Information Modeling (BIM) to offer valuable project insights including lessons learned, value generation, and continuous improvement [19], [20]. The convergence of Lean principles with Industry 4.0 within a maturity model context appears to be scarcely explored. To the best of our knowledge, [23] is the only article, published in 2023, to introduce a digital Lean maturity model in this intersection.

Exploring further both paradigms, [23] proposed a maturity model based on Lean and Industry 4.0 synergy. The model consists of four key elements: strategic pillars, perspectives dimensions and maturity levels. The model comprises two main strategic pillars: Lean and Industry 4.0. These pillars serve as the fundamental concepts guiding the framework. Additionally, there are three key perspectives termed as the "Smart" components, which include processes, people, and products.

These perspectives offer a holistic view of the organizational landscape. Moreover, the model defines maturity levels for both the strategic pillars, termed as strategic maturity levels, and the perspectives, termed as "Smart" maturity levels. This categorization aids in assessing the organization's progression within each aspect of the framework.

To our knowledge, all existing maturity models (Lean, industry 4.0 and Lean 4.0 models) use an assessment matrix to evaluate current maturity levels and rely solely on these levels to identify areas for improvement. Our machine learning model bridges this gap by leveraging similar use cases. Instead of offering generic recommendations, our model provides tailored roadmaps specific to each company's characteristics to enhance their maturity levels. Put simply, we learn from others' experiences to avoid errors and benefit from their successes. Instead of just using formulas and rules, we look at real-life examples to create a roadmap and help companies on their journey.

III. DATASET AND METHODOLOGY

The flowchart depicting our study's methodology is illustrated in Figure 1. Initially, the process involves obtaining a reliable database.

Due to the challenge of accessing databases containing real-world use cases of Lean 4.0 projects, primarily due to confidentiality concerns, we utilized a limited dataset as provided in reference [24]. The database consists of 19 companies and 53 attributes. Nevertheless, given the fact that we explore Machine Learning, a sufficient data amount should be provided to ensure obtaining an efficient model. Thereby, we use the segmentation-based data augmentation technique to increase the size of the data from 19 to 274 companies.

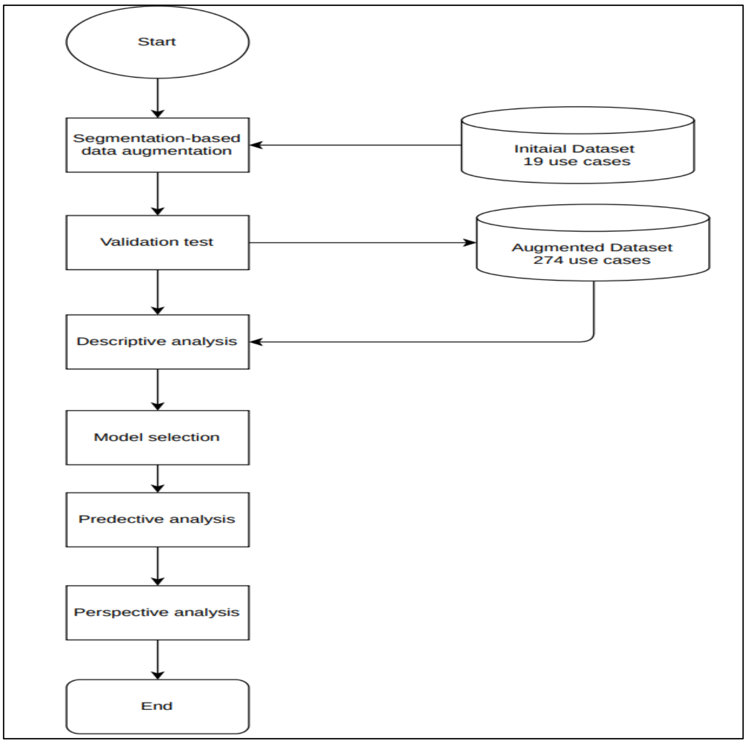

Figure 1: ML Workflow Chart.
Source: Authors, (2024).

Then, we verify the validity of our augmented data by calculating the percentage of maintained relationships between attributes in the generated data before incorporating it into our analysis. Before proceeding with predictive and perspective analysis, it is essential to select an appropriate model to ensure optimal results and insightful outcomes. This involved choosing the most suitable machine learning model that would effectively extract valuable insights from the data. This analysis process involves three key stages:

1.Descriptive Analysis: Here, we examine the correlations between different attributes and investigate the factors that have the most significant influence on the maturity of Industry 4.0.

2.Predictive Modeling: We utilize similar cases to predict the maturity level, leveraging predictive analytics techniques.

3.Perspective Enhancement: This stage involves offering customized and personalized recommendations aimed at enhancing the maturity level based on the insights gained from the analysis.

The authors of [24] focused on the difference between the companies with Low Lean maturity level and those with high Lean maturity level in the adoption of I4.0 technologies. Leveraging the same dataset, we aim to construct a maturity model, manifested as a machine learning (ML) model, designed to aid companies in forecasting their maturity levels and furnishing tailored recommendations based on analogous cases.

The contribution consists of: 1) Expanding the database using augmented data techniques to use in analyze phase; 2) Analyze the factors influencing maturity levels using machine learning tools; 3) optimize parameters of the model and 4) propose recommendations using decision tree.

III.1. DETAILS OF THE DATASET

The database (BD) presented in [24] consists of 19 use cases and 53 attributes. The companies involved in this study are confined to just two sectors: Machinery (M) and Metal Products (MP). These 19 companies fall into one of two categories: those with a Lean maturity level zero (0), indicating a low level of Lean maturity, and those with a Lean maturity level one (1), indicating a high level of Lean maturity.

These attributes are classified into four dimensions: 1) Industry 4.0 technologies and maturity level; 2) Targeted area of industry 4.0 investments; 3) purpose and expected performance of industry 4.0 investments and 4) magnitude of industry 4.0 investments. Each dimension has multiple sub-dimensions which are the variables of our model.

In this database, the authors consider 4 maturity levels (table 1) to evaluate six I4.0 technologies: 1) IoT; 2) Industrial analytics; 3) Advanced human–machine interface; 4) Cloud manufacturing; 5) Additive manufacturing and 6) Advanced automation. These technologies serve as the sub-dimensions falling under the primary dimension of "Industry 4.0 Technologies and Maturity Level."

Table 1: Maturity levels.

Level 0	Not Applicable
Level 1	Monitoring
Level 2	Control
Level 3	Control / Optimization Autonomy
Level 4	Optimization Autonomy

Source: [24]0.

In our study, we aim to assess the overall maturity level of the company by considering these different technologies. Therefore, we propose to replace the maturity level of each technology by an overall maturity level. This level is the weighted average of the maturity levels of the 6 technologies. The weights of each technology were defined in [25] using the Analytic Hierarchy Process (AHP) method. The weighting factors were assigned as follows:

- IoT Maturity level: 15.9%
- Industrial analytics Maturity level: 19.4%
- Advanced human–machine interface Maturity level: 40.2%
- Cloud manufacturing Maturity level: 11.7%
- Additive manufacturing Maturity level: 3.7%
- Advanced automation Maturity level: 3.2%
- Simulation Maturity level: 5.3 %

Here, we shall point out that these weights have been validated by calculating the consistency index of the related pay-off judgmental matrix. The obtained ratio is 0.02 which ensures that judgments established for computing these weights are coherent since the default risk is inferior to 0.1 (a conventional threshold that is typically considered).

Due to the absence of any mention of the "Simulation" technology in the responses gathered by [24] during their survey on technology investments, we propose excluding it from our analysis. In other words, since "Simulation" did not feature in the survey data, its value will be considered as 0. Thus, it will consequently be omitted from the equation used to determine the 'Overall Maturity Level'. The degree of Industry 4.0 development depends on the level of advancement in Industry 4.0 technologies, as shown by equation (1):

*data ['Overall Maturity Level'] = 0.159 * data ['IOT-Maturity Level'] + 0.194 * data ['Industrial analytics - Maturity level'] + 0.402 * data ['Advanced human–machine interface - Maturity level'] + 0.117 * data ['Cloud manufacturing - Maturity level'] + 0.037 * data ['Additive manufacturing - Maturity level'] + 0.032 * data ['Advanced automation - Maturity level']* *(1)*

To summarize, we have created a new column, **['Overall Maturity Level']**, in the **data** dataset. This new column replaces the six individual columns representing maturity levels in the six specified technologies. The value of **['Overall Maturity Level']** is calculated as the weighted sum of the maturity levels across these six technologies, as previously described.

III.2. DATA-SEGMENTATION BASED DATA AUGMENTATION

The dataset suggested in [24] is insufficiently large to construct a highly effective machine learning model. It only consists of nineteen use cases. To address this limitation, we can employ data augmentation techniques to generate additional case studies, thereby facilitating the development of a more accurate model.

The used approach customizes data augmentation to distinct data segments. It is fostering a more nuanced and effective learning process for machine learning models dealing with tabular data. It is valuable for tabular data due to its ability to preserve relationships and facilitate feature-specific augmentation. This technique is implemented through the following steps:

1. Hierarchical Clustering:

● Use hierarchical clustering to group data points into clusters based on their similarity.

● The optimal number of clusters is determined using the elbow method (in this case, it's set to 3).

2. Assigning Clusters:

● Assign each data point to a cluster based on the results of hierarchical clustering.

● This is done using the Agglomerative Clustering algorithm which is a bottom-up approach that starts by considering each data point as a single cluster and then progressively merges the closest pairs of clusters until only one cluster remains.

3. Data Augmentation Function:

● Define a function named augment_data that takes two arguments: data (the original dataset) and num_samples (the number of augmented samples to generate).
● Create a copy of the original data to store augmented data.

4. Loop Through Columns:

● Iterate over each column in the dataset except for the target column and the cluster column.

5. Augmentation Loop:

● For each column, loop through num_samples times to augment data.
● Randomly select a cluster from the clusters present in the dataset.
● Extract the data points belonging to the selected cluster.
● Randomly select one data point from the cluster.
● Perturb the value of the selected column in the chosen data point by randomly choosing a value from the unique values of that column in the original dataset.
● Append the augmented data point to the augmented dataset.

6. Return Augmented Data:

● Return the augmented dataset.

Prior to proceeding to data augmentation, we cluster it with a view to obtain homogeneous partitions. This would be of interest to apply the data augmentation safely by preventing high intra-variance. To do so, we referred to the Hierarchical clustering. According to [26], Hierarchical clustering is more suitable for categorical data. It also provides a hierarchical structure referred to as dendrogram that illustrates the relationships and the internal structure between clusters. This algorithm does not require the specification of the number of clusters in advance and can handle clusters of various shapes and sizes.

As depicted in Figure 2, the dendrogram displays various clusters of companies based on their similarities. The vertical lines linking companies are proportional to the level of similarity between the connected companies. The shorter the line, the more similar the companies are. For instance, from Figure 2, we observe that companies 17 and 5 are the most similar.

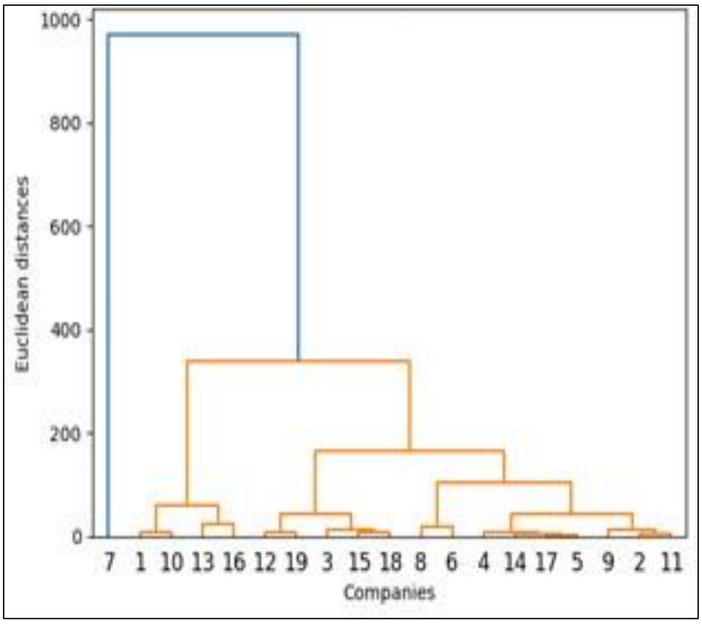

Figure 2: Dendrogramme des Entreprises : Analyse de Similarité et Clustering Hiérarchique.
Source: Authors, (2024).

We determined the number of clusters for our data using a method called the elbow method. This method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and looking for the point where the rate of decrease in WCSS slows down, forming an "elbow" shape in the plot. In our analysis, we observed that the elbow point occurred when we had three clusters. The next step consists of applying Data augmentation to each cluster.
The subsequent step involves applying Data augmentation to each cluster. The augmented data size amounts to 274.

Proposed a multivariate relationship analysis method to assess the fidelity of generated data compared to real data. This technique involves comparing Pairwise Pearson Correlation matrices between the real and generated data through heatmaps. A heatmap is a graphical representation of data where individual values are displayed using colors [27].

It is common practice to visualize the intensity or distribution of values across two dimensions, such as rows and columns in a matrix. By calculating the differences in correlations between the two datasets, the percentage of maintained relationships in the generated data is determined. If this percentage exceeds 0.6, the approach is classified as "Excellent"; between 0.4 and 0.6, as "Good"; and below 0.4, as "Poor".

In our case, the Percentage of numerical relationships maintained in generated data is 0.43 which means that our data augmentation approach is qualified as Good. This suggests that the generated data adequately preserves a substantial portion of the multivariate relationships observed in the real data.

III.3. UTILIZED MODELS AND TECHNIQUES

As shown in Figure 1, our analysis starts with descriptive analysis, then predictive analysis and finally perspective analysis. Each step from this process requires different techniques. For descriptive analysis, we selected Multiple Component Analysis (MCA) for multidimensional analysis and Heatmap.

By using these methods, we attempt to meet two objectives. First, identify pertinent features to the maturity level analysis by retaining only one example of each highly correlated pair of features. This would be of interest to refine the data to be fed to the machine learning model and prevent any overfitting issue. Second, identify the linkages between features and the maturity levels of companies. This allows to emphasize improvements scopes for the least performing. ML models enable us to achieve predictive analysis objectives.
In our scenario, we've opted for Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM) to conduct our predictive analysis on forecasting the maturity level of a company. The selection of these specific ML models stems from their distinct strengths and suitability for the task at hand.
1. Decision Trees (DT): is chosen for its simplicity and interpretability. It's adept at handling categorical and numerical data, making it suitable for diverse datasets. The technical steps involved in building a decision tree include:
● Selecting the best attribute to split the data at each node, usually based on metrics like Gini impurity or entropy.
● Recursively partitioning the data based on these attributes until a stopping criterion is met, such as reaching a maximum depth or purity threshold.
2. Random Forests (RF): is an ensemble learning method that combines multiple decision trees to improve predictive performance and reduce overfitting. It's particularly useful when

dealing with high-dimensional data or datasets with a large number of features. The technical steps of RF involve:

● Building multiple decision trees on random subsets of the training data (bootstrapping).

● Aggregating the predictions of individual trees through voting or averaging to make the final prediction.

3. Support Vector Machines (SVM): is chosen for its effectiveness in handling both linear and non-linear classification tasks. It works by finding the hyperplane that best separates different classes while maximizing the margin between them. The technical steps of SVM include:

● Mapping the input data into a higher-dimensional space using a kernel function.

● Finding the optimal hyperplane that separates the classes with the maximum margin or minimizing classification errors.

By employing these models, we aim to leverage their unique capabilities to accurately predict the maturity level of companies. This predictive capability is crucial for strategic planning, as it empowers proactive decision-making to optimize processes and enhance overall maturity level.

IV. RESULTS AND DISCUSSIONS

IV.1. DESCRIPTIVE ANALYSIS

It is important to note that our dataset contains categorical variables. Therefore, we have selected Multiple Correspondence Analysis (MCA) technique which is a useful technique for multidimensional analysis of categorical data [28]. This technique enables the condensation of a group of categorical variables into a limited set of independent variables known as principal components. These components optimally summarize the data given in [28].

Now, let's explore the insights from the MCA graph (Figure 3):

a) Companies on the Upper Right Side of the Plot (Near Company 11):

The companies clustered closely together on the graph share remarkably similar characteristics, especially in their advanced industrial analytics, advanced automation, IoT, and human-machine interface maturity levels. What stands out is that they all show high levels of maturity in these areas, along with a significant use of extensive long-term interventions. This observation aligns with the conclusions drawn in reference [24]. In fact, it suggests that companies focusing on such interventions tend to have lower lean maturity levels but higher maturity in Industry 4.0. Therefore, this alignment strengthens the conclusion that these companies indeed demonstrate heightened maturity in Industry 4.0 technologies, backed by supporting evidence.

b) Companies on the Upper Left Side of the Plot (Close to Each Other):

This particular group of companies stands out as a cohesive cluster, evident from their close proximity on the individual-variable graph. Their distinguishing features include a preference for localized medium-term interventions, substantial investments in Industry 4.0 logistics, a modest maturity level in advanced automation (Level 0), and a leaning towards Lean principles. Our analysis suggests a correlation where companies with elevated Lean maturity prioritize logistics investments while demonstrating lower maturity levels in advanced automation.

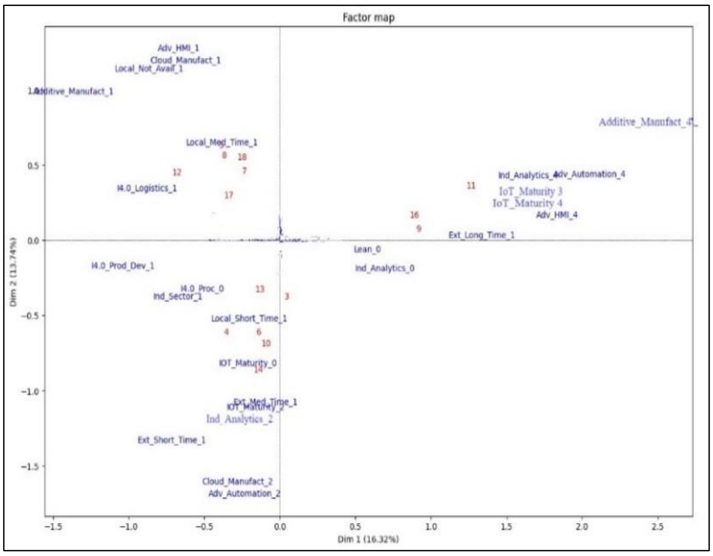

Figure 3: MCA Factor Map of Key Variables.
Source: Authors, (2024).

Now that we've explored the insights from the MCA graph, let's transition to discussing the heatmap. Transitioning from the Multiple Correspondence Analysis (MCA) graph to the heatmap serves a crucial purpose in our analysis. While the MCA provides an overview of relationships between categorical variables, the heatmap allows us to delve deeper into specific interactions and correlations. By visualizing the data in a heatmap, we can identify patterns, dependencies, and potential areas of interest. [29] describes the Heatmap (correlation matrix) as a necessary input for others who may wish to reproduce (and confirm) a study's results, as well as perform secondary analysis.

In this study, we employ heat maps as a powerful tool for conducting descriptive analysis of our dataset. Heat maps offer a visually intuitive means of summarizing and exploring the distribution of quantitative variables within our data, allowing us to identify prominent patterns and trends. By representing data values with varying colors, heat maps provide immediate insights into the relative magnitudes and spatial relationships of different variables or observations.

The heat map graph of Figure 4 provides a visual representation of the relationships between different variables in our dataset. These correlations offer several interpretations and insights into the factors influencing the overall maturity level of Industry 4.0 technologies. We summarize this into four factors:

a) 'Overall Maturity level' – Industry 4.0 inves - Assembly: 0.57:

The strong positive correlation suggests that organizations with higher levels of investment in assembly processes tend to have higher overall maturity levels in their Industry 4.0 technologies. This implies that a strategic focus on optimizing assembly operations through technology investments contributes significantly to the overall advancement of Industry 4.0 capabilities.

b) 'Overall Maturity level' - Quality -Governmental incentives: 0.46 / Overall Maturity level' - Productivity - Governmental incentives: 0.42:

The moderate positive correlations indicate that governmental support in the form of incentives for quality and productivity improvements is associated with higher overall maturity levels in Industry 4.0 technologies. Organizations that leverage such incentives may have access to resources or initiatives that facilitate the adoption of advanced technologies and practices, leading to enhanced quality and productivity outcomes.

c) 'Overall Maturity level' - Extensive interv - Long Time: 0.42:

The correlation of 0.42 between the overall maturity level and extensive interventions over a long time suggests that companies investing in long-term interventions tend to have higher maturity levels in Industry 4.0. This aligns with findings from [24], where such companies demonstrated higher Industry 4.0 maturity despite lower Lean maturity levels. In essence, sustained long-term efforts in adopting Industry 4.0 technologies positively impact overall maturity levels, indicating a strategic focus on digitalization and technological advancement over time.

d) 'Overall Maturity level' – Flexibility -Improving waste detection: 0.01 / 'Overall Maturity level' - Industry 4.0 invest - processing: 0.01:

The modest correlation between the "Overall Maturity Level" and the flexibility-improving waste detection suggests that companies prioritizing flexibility enhancements and waste detection in their industry 4.0

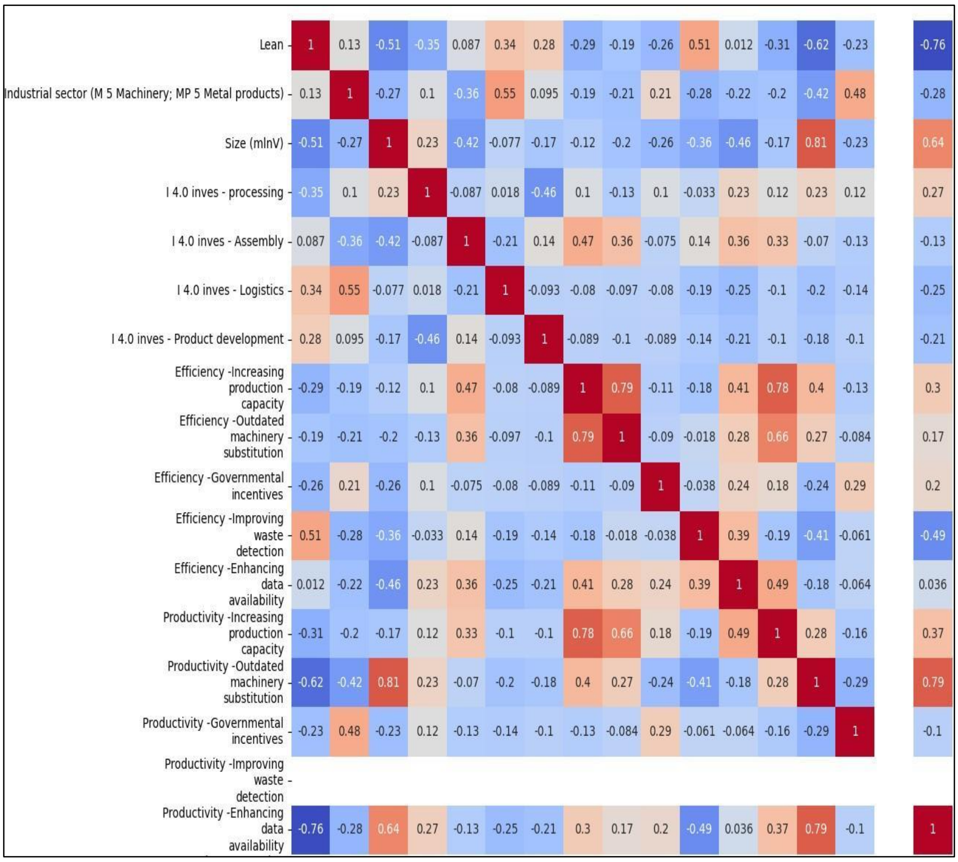

Figure 4: Partial Heat map Focus (correlation matrix).
Source: Authors (2024).

initiatives might not necessarily exhibit a high Industry 4.0 maturity level. These factors seem to have limited influence on the overall maturity level. Similarly, the weak correlation between investment in processing and overall maturity level implies that decisions regarding processing investment may not significantly impact Industry 4.0 maturity.

While analyzing the data, we observe that the “Lean” variable does not prominently appear among the influencing factors for the overall maturity level of Industry 4.0. Companies with Low Lean maturity level and High maturity level approach industry 4.0 differently. Despite these variations, the overall maturity level of Industry 4.0 remains consistent. In other words, although these companies implement Industry 4.0 differently, the overall maturity level shows only minor variations.

In summary, while the “Lean” factor does not significantly influence the overall maturity level of Industry 4.0, it plays a crucial role in shaping the implementation approach. Companies may adopt different strategies, yet the overall maturity level remains relatively consistent across both low and high Lean maturity companies.

IV.2. MODEL SELECTION

All of these models we used have the same phases. They first go through feature engineering which consists of feature selection and data cleaning. Following feature engineering, the models undergo training using labeled data to learn patterns and relationships between features and the target variable. During the training phase, model parameters are optimized to minimize a predefined loss function. Once trained, the models are evaluated using validation data to assess their performance and fine-tune hyperparameters if necessary. Finally, the models are deployed to make predictions on unseen data, providing insights into the maturity level of companies based on their features.

We chose Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2) score to evaluate the performance of our models because each metric offers unique insights into different aspects of model performance. MSE provides a comprehensive measure by penalizing larger errors more heavily, making it suitable for scenarios where large errors are critical. MAE, on the other hand, treats all errors equally, offering a simpler and more interpretable measure of performance that is less sensitive to outliers. Lastly, R2 score allows us to assess the goodness of fit of the model by indicating how well the independent variables explain the variability of the dependent variable. Together, these metrics provide a well-rounded evaluation of model accuracy, robustness to outliers, and overall fit to the data.

The MSE calculates the average of the squared errors (differences between predicted and actual values) [30]. The MAE represents the average absolute difference between the predicted values and the actual target values. Lower MSE and MAE (closer to zero) indicate better performing model.

R2 score assesses how well our model performs by measuring the proportion of explained variance. As far as the R2_score value gets near to one a better model performance is retrieved.

Table 2: Statistical performances of compared models.

AI model Metrics	Decision Tree	Random Forest	SVM
MSE	0.033	0.036	0.033
MAE	0.062	0.089	0.111
R2_score	0.863	0.851	0.863

Source: Authors, (2024).

As may be noticed from table 2, the Decision Tree model stands out as the most promising, demonstrating superior performance with an MSE of 0.033, MAE of 0.062, and an R2 score of 0.863. We note that, beyond its exemplary performance in this context, the Decision Tree model offers additional advantages that further justify its selection. Decision trees are inherently interpretable, allowing for straightforward visualization and understanding of decision-making processes. They are also computationally efficient and less sensitive to outliers compared to other models like SVM. Moreover, Decision Trees naturally handle feature interactions and nonlinear relationships, making them well-suited for datasets with complex structures. Hence, given its superior performance and inherent advantages, the Decision Tree model emerges as the optimal choice for this task, underscoring its versatility and efficacy in practical machine learning applications.

A Decision Tree is a machine learning technique that helps categorize (classification) or forecast values (regression). It divides data into sections based on their traits and assigns a label or predicts a value for each section. Key parameters for Decision Trees include:

1. Split Criteria: Defines how to assess the quality of a data split. Common choices include "Gini impurity" and "information gain."
2. Maximum Depth: Limits how deep the tree can grow, preventing excessive detail and reducing overfitting.
3. Minimum Split Samples: Sets the minimum number of data points needed to divide a node, helping to prevent splitting based on too little information.
4. Minimum Samples Leaf: Sets the minimum number of samples required for a node to be considered a leaf node (terminal node without any further splits).

With our Decision Tree model, we've carefully chosen parameters using grid search to optimize its performance. For the "min samples leaf" parameter, we initiated the numerical list in the grid search at 8. This decision reflects our commitment to creating a model that not only accurately captures patterns in the data but also ensures robustness, interpretability and generalization. By setting a minimum number of samples for each leaf node, we're encouraging the model to make decisions based on larger subsets of the data. This helps in extracting meaningful insights and interpretations from the decision tree, as we're analyzing patterns based on groups of data points rather than individual instances.

The grid search examines many combinations of params and provides us with the optimal combination which refers to the best performing model (Table 3).

Table 3: Best Hyperparameters of decision tree using Grid Search.

Split Criteria	Max_Depth	Min_Samples_Leaf	Min_Samples_Split
Gini impurity	None	8	2

Source: Authors, (2024).

In summary, Random Forest and SVM are only useful for descriptive and predictive analysis. However, the Decision Tree allows us to make the three (3) analysis (descriptive, predictive, and perspective). Besides, the Decision Tree is performing better than the other models. As a result, the Decision Tree is the most suitable choice for our analysis.

IV.3. PREDICTIVE ANALYSIS

After analyzing the key features, understanding the correlations and patterns in our data (Descriptive analysis), we can use these features for predicting the current maturity level of the company.

Figure 5 shows our Decision Tree model. Each node represents a question, and each branch corresponds to a response to that question [31]. Conceptually, we can view each node as a decision point. By traversing the tree and answering the questions at each node until reaching the last node, also called a leaf node, we can predict the maturity level based on the path followed.

To illustrate the application of the model, let's take the example of a case study (red circle in Figure 5). We start from the top of the decision tree (Figure 5). We observe that companies obeying the condition of the parent node (1st node) emphasize investments in Assembly. We move forward to the next node (2nd node), we go with companies that do not satisfy this condition. That means, companies who are not expecting Industry 4.0 performance improvements in Quality while also prioritizing enhancing data availability as an investment driver. For the next node (3rd node) we choose companies whose size exceeds 111. We reach the 4th node and observe that companies who are not adopting extensive interventions for long term tend to have an overall Industry 4.0 maturity level around 1.176 (5th node = Leaf node). By following this decision path (red circle), we can predict the maturity level for companies sharing these same attributes and decisions, which in this case corresponds to 1.176.

IV.4. PERSPECTIVE ANALYSIS

Our goal is to increase the maturity level. By examining the nodes following the leaf we can pinpoint actionable decisions contributing to maturity enhancement. It's crucial to distinguish between contextual variables like size or sector and actionable decisions, as changing contextual variables might not be feasible. By following this methodology, companies can pinpoint precise interventions necessary for advancing maturity levels, thereby leveraging the Decision Tree as a practical recommendation tool for organizational improvement.

Taking the example of the company associated with the leaf node displaying an overall maturity level of 1.176. By bringing together all companies that meet the same conditions present in the parent nodes of this sheet, it is likely that they share a similar maturity level because they conform to the same characteristics. Based on our model and considering similar cases, it is recommended that these companies emphasize extensive interventions for long term to improve their industry 4.0 maturity level. This recommendation is substantiated by findings from [7]—[24], who suggest that companies with low Lean maturity levels often require substantial interventions to foster significant improvements.

By adopting this strategy, there is a high probability, based on other similar cases, that this company can increase their maturity level from 1.176 to 1.381.

Furthermore, our analysis reveals a noteworthy observation: companies that neglect localized interventions exhibit a relatively lower maturity level of 0.755. However, by simply adopting localized interventions, this maturity level can be significantly enhanced to 0.876. This finding underscores the pivotal role of localized interventions in driving organizational maturity within the context of Industry 4.0. According to [7]—[24], companies that prioritize localized interventions typically demonstrate high levels of Lean maturity. This correlation suggests that by emphasizing localized interventions, companies not only enhance their Industry 4.0 maturity level but also align with established principles of Lean methodology. Therefore, integrating localized interventions into organizational strategies not only facilitates maturity enhancement but also fosters a culture of lean thinking and continuous improvement. Accordingly, it enables companies to position themselves for sustained success in the rapidly evolving landscape of Industry 4.0.

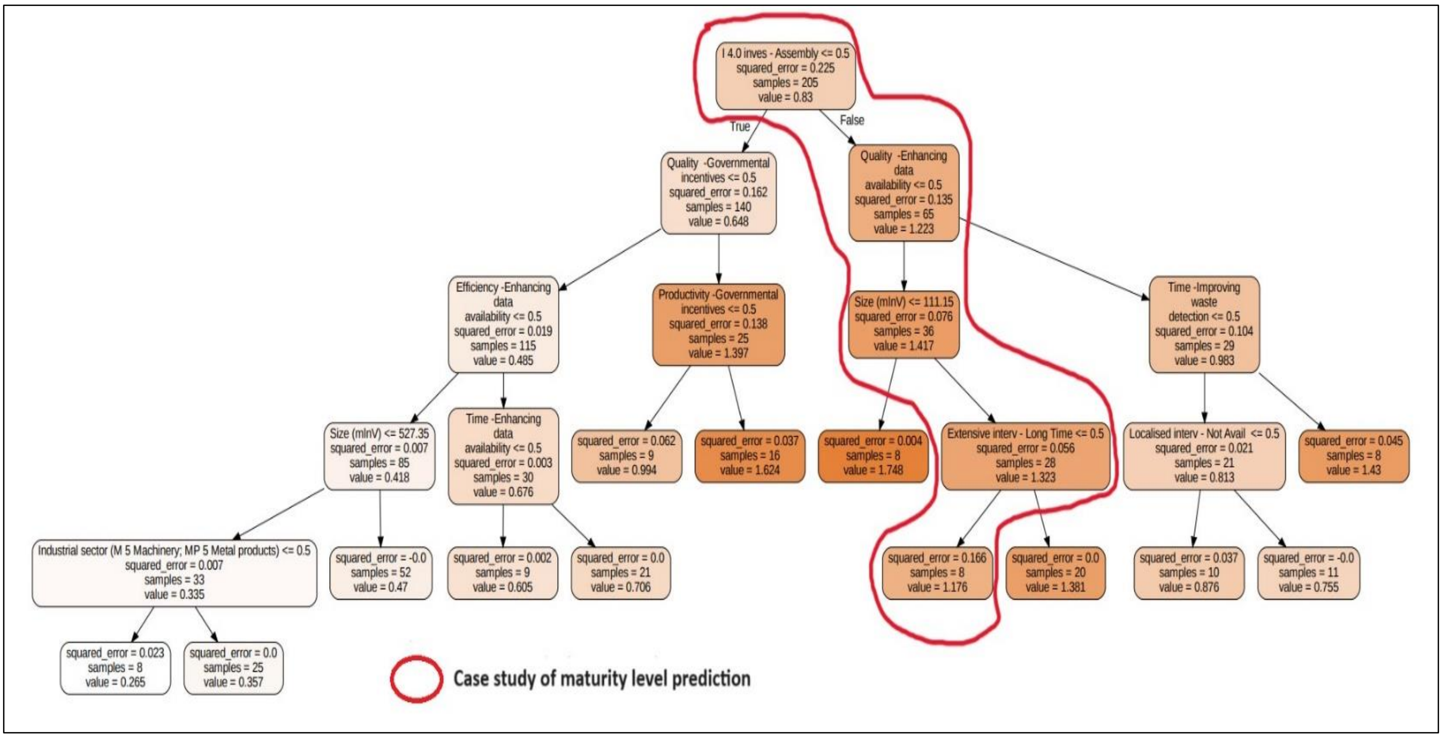

Figure 5: Decision Tree based ML model for maturity prediction and enhancement.
Source: Authors, (2024).

Besides, these companies can also follow a second path “Time – improve waste detection” which can significantly increase their overall maturity level from 0.755 to 1.43.

It indicates a crucial aspect influencing the maturity level of Industry 4.0 within the context of waste detection improvement. The presence of this column suggests that dedicating resources and efforts towards enhancing waste detection processes is pivotal for advancing Industry 4.0 maturity. This linkage between time-related performance improvement and waste detection aligns closely with Lean principles.

In Lean methodology, waste detection and reduction are fundamental pillars aimed at maximizing value and efficiency while minimizing resources and time. By focusing on improving waste detection as an Industry 4.0 investment driver, organizations align with Lean principles by identifying and eliminating non-value-adding activities or processes. The "Time – improve waste detection" column signifies an acknowledgment of the importance of time efficiency in waste detection efforts. Lean principles emphasize the elimination of waste to streamline processes and reduce lead times, thereby enhancing overall operational performance.

To create a straightforward plan of suggested actions, we looked at the decision tree (Figure 5) and collected possible recommendations along with the conditions that trigger them, putting everything together in a single table (Table 4).

Table 4 outlines the recommendations (Actions A1 to A6) corresponding to the conditions presented in the top section of the table (C1 to C12). It offers a customized roadmap of suggested actions for each company based on its features (conditions). In the table, 'Y' indicates that a condition is met by the company, 'N' indicates the opposite, and blank cells signify that the condition has no effect on the given actions. The recommended actions for each set of rules are denoted by 'X'.

Table 4: Recommendation table.

Conditions		Rules				
		1	**2**	**3**	**4**	**5**
C1	Are there Industry 4.0 investments in Assembly?	Y	Y	Y	N	N
C2	Is Quality targeted for Industry 4.0 performance improvement, with governmental incentives as the investment driver?				Y	N
C3	Is Quality targeted for Industry 4.0 performance improvement, with Enhancing data availability as the investment driver?	Y	Y	N		
C3	Is Time targeted for Industry 4.0 performance improvement, with Improving waste detection as the investment driver?	N	N			
C4	Is Time targeted for Industry 4.0 performance improvement, with Improving waste detection as the investment driver?			Y		
C5	Is company size <=111.15				N	
C6	Is Productivity targeted for Industry 4.0 performance improvement, with governmental incentives as the investment driver?					Y
C7	Is Efficiency targeted for Industry 4.0 performance improvement, with Enhancing data availability as the investment driver?	N	Y			
C8	Does the company refrain from adopting localized interventions?			N		
C9	Is the company implementing Extensive interventions for the long term?					
C10	Is Time targeted for Industry 4.0 performance improvement, with Enhancing data availability as the investment driver?					N
C11	Is company size <=527.35					
C12	Is the industrial sector Metal Products?					
Actions		**1**	**2**	**3**	**4**	**5**
A1	Target Industry 4.0 performance improvement using Time, with Improving waste detection as the investment focus.	X	X			
A2	Implement localized interventions.	X				
A3	Implement extensive interventions for long term.			X		
A4	Target Industry 4.0 performance improvement using Productivity, with gouvermental incentives as the investment focus.				X	
A5	Target Industry 4.0 performance improvement using Time, with Enhancing data availability as the investment focus.					X
A6	Target Industry 4.0 performance improvement using Quality, with gouvermental incentives as the investment focus.					X

Legend: Y: Yes, **Empty case:** Not Applicable, **N:** No, **X:** Recommended Action.
Source: Authors, (2024).

V. CONCLUSIONS

This paper deals with the factors influencing the industry 4.0 maturity level. To this end, we study the impact of lean maturity on Industry 4.0 maturity level. Besides, we explore maturity models and recommendations for digital Lean initiatives.

Thus, the integration of machine learning (ML) models, particularly the Decision Tree algorithm offers significant utility. Primarily, it is noticed in the context of predicting and enhancing maturity levels. Our study showcases the effectiveness of ML models in accurately predicting Industry 4.0 maturity levels based on company characteristics. This provides actionable recommendations for improvement. By leveraging the Decision Tree model, we were able to forecast maturity levels. In addition, we identify key factors influencing organizational maturity and suggest targeted interventions for enhancement. This approach serves as a valuable tool within the framework of maturity models, enabling organizations to systematically assess their current state, identify areas for development, and implement tailored strategies for advancement. Moreover, the interpretability and transparency of Decision Trees make them particularly suitable for decision-making processes in industry. They foster a deeper understanding of the underlying factors driving maturity levels. Overall, the utilization of Decision Trees, as decision support tools, offers a practical and efficient means of navigating the complexities of organizational maturity. Ultimately, Decision Trees facilitate informed decision-making and drive continuous improvement in industry practices.

The current study was initially confined to a relatively small data set comprising 19 real-world use cases and focused only on two sector activities, namely Machinery and Metal Products. Future research could involve expanding the dataset to encompass a more extensive range of sectors and varying sizes of companies. Additionally, there exists an opportunity to explore alternative targets within the same database to inform strategic decision-making processes. For instance, the columns representing targeted areas, such as processing and assembly, could be individually designated as targets to assess the feasibility. This enables the evaluation of potential benefits of investing in each area. Furthermore, by modifying the target variable in our model, insights into expected gains metrics like Return on Investment (ROI) could be achieved through survey-based approaches. This adaptation would facilitate a more comprehensive evaluation of the potential outcomes and benefits associated with different investment strategies.
Example:

VI. AUTHOR’S CONTRIBUTION

Conceptualization: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H’Mida, Ahmed Nait Sidi Moh.
Methodology: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H’Mida, Ahmed Nait Sidi Moh.
Investigation: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H’Mida, Ahmed Nait Sidi Moh.
Discussion of results: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H’Mida, Ahmed Nait Sidi Moh.

Writing – Original Draft: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H'Mida, Ahmed Nait Sidi Moh.
Writing – Review and Editing: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H'Mida, Ahmed Nait Sidi Moh.
Resources: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H'Mida, Ahmed Nait Sidi Moh.
Supervision: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H'Mida, Ahmed Nait Sidi Moh.
Approval of the final text: Oussama Ben Ali, Sondes Hammami, Marwa Hasni, Fehmi H'Mida, Ahmed Nait Sidi Moh.

VII. REFERENCES

[1] Weyer, S., Schmitt, M., Ohmer, M., & Gorecky, D. (2015). Towards Industry 4.0-Standardization as the crucial challenge for highly modular, multi-vendor production systems. Ifac-Papersonline, 48(3), 579-584. https://doi.org/10.1016/j.ifacol.2015.06.143

[2] Rosin, F., Forget, P., Lamouri, S., & Pellerin, R. (2019). Industrie 4.0 et Lean : Hiérarchie, amélioration ou complémentarité ? Conference: ClGl QUALITA 2019, "Le numérique et l'intelligence industrielle: défis et opportunités".

[3] Rossini, M., Cifone, F. D., Kassem, B., Costa, F., & Portioli-Staudacher, A. (2021). Being lean: how to shape digital transformation in the manufacturing sector. Journal of Manufacturing Technology. https://doi.org/10.1108/JMTM-12-2020-0467

[4] Rossini, M., Costa, F., Tortorella, G. L., & Portioli-Staudacher, A. (2019). The interrelation between Industry 4.0 and lean production: an empirical study on European manufacturers. The International Journal of Advanced Manufacturing Technology, 102, 3963-3976. https://doi.org/10.1007/s00170-019-03441-7

[5] Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. Engineering structures, 171, 170-189. http://dx.doi.org/10.1016/j.engstruct.2018.05.084

[6] Iversen, J., Nielsen, P. A., & Norbjerg, J. (1999). Situated assessment of problems in software development. ACM SIGMIS Database: the Database for Advances in Information Systems, 30(2), 66-81. http://dx.doi.org/10.1145/383371.383376

[7] Wendler, R. (2012). The maturity of maturity model research: A systematic mapping study. Information and software technology, 54(12), 1317-1339. https://doi.org/10.1016/j.infsof.2012.07.007

[8] Pöppelbuß, Jens, and Maximilian Röglinger. (2011) "What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management.". *19*th European Conference on Information Systems, ECIS 2011

[9] Rummler, G. A., & Brache, A. P. (2012). Improving performance: How to manage the white space on the organization chart. John Wiley & Sons.

[10] Amaral, A., & Peças, P. (2021). A framework for assessing manufacturing SMEs Industry 4.0 maturity. Applied Sciences, 11(13), 6127. http://dx.doi.org/10.3390/app11136127

[11] Cheshmberah, Mohsen, and Safoura Beheshtikia (2020) "Supply chain management maturity: An all-encompassing literature review on models, dimensions and approaches." LogForum 16.1: 103-116. https://doi.org/10.17270/J.LOG.2020.377

[12] Maier, Anja & Moultrie, James & Clarkson, P. (2009). Developing maturity grids for assessing organizational capabilities: Practitioner guide. 1-29.

[13] Alcácer, V., Rodrigues, J., Carvalho, H., & Cruz-Machado, V. (2022). Industry 4.0 maturity follow-up inside an internal value chain: a case study. The International Journal of Advanced Manufacturing Technology, 1-12. http://dx.doi.org/10.21203/rs.3.rs-431651/v1

[14] Leyh C, Schäffer T, Bley K, Forstenhäusler S (2016) SIMMI 4.0 – a maturity model for classifying the enterprise-wide IT and software landscape focusing on Industry 4.0. Proceedings of the Federated Conference on Computer Science and Information Systems. 8:1297–1302. https:// doi. org/ 10. 15439/ 2016F 478.

[15] De Carolis A, Macchi M, Negri E, Terzi S (2017) A maturity model for assessing the digital readiness of manufacturing companies, in advances in production management systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing, Cham 13–20. https:// doi. org/ 10. 1007/ 978-3- 319-66923- 6_2

[16] Rajnai Z, Kocsis I (2018) Assessing industry 4.0 readiness of enterprises. 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI) 225–230. https:// doi. org/ 10.1109/ SAMI. 2018. 83248 44.

[17] Rockwell Automation (2014) The connected enterprise maturity model. https:// literature. Rockwell au tomation. com/ idc/ groups/ literature/ documents/ wp/ cie-wp002_- en-p. pdf

[18] Jayanetti, J. K. D. D. T., Perera, B. A. K. S., Waidyasekara, K. G. A. S., & Siriwardena, M. (2023). Critical Analysis of Lean Construction Maturity Models: A Systematic Literature Review. Buildings, 13(6), 1508. https://doi.org/10.3390/buildings13061508

[19] Demirdöğen, G., Diren, N. S., Aladağ, H., & Işık, Z. (2021). Lean based maturity framework integrating value, BIM and big data analytics: Evidence from AEC industry. Sustainability, 13(18), 10029. https://doi.org/10.3390/su131810029

[20] Mollasalehi, S., Aboumoemen, A. A., Rathnayake, A., Fleming, A. J., & Underwood, J. (2018). Development of an integrated BIM and lean maturity model. Conference: 26th Annual Conference of the International Group for Lean Construction http://dx.doi.org/10.24928/2018/0507

[21] Nightingale, D. J., & Mize, J. H. (2002). Development of a lean enterprise transformation maturity model. Information Knowledge Systems Management, 3(1), 15-30.

[22] Maasouman, M. A., & Demirli, K. (2016). Development of a lean maturity model for operational level planning. The International Journal of Advanced Manufacturing Technology, 83, 1171-1188. http://dx.doi.org/10.1007/s00170-015-7513-4

[23] Treviño-Elizondo, B. L., García-Reyes, H., & Peimbert-García, R. E. (2023). A maturity model to become a Smart Organization based on lean and Industry 4.0 synergy. Sustainability, 15(17), 13151. https://doi.org/10.3390/su151713151

[24] Rossini, M., Cifone, F. D., Kassem, B., Costa, F., & Portioli-Staudacher, A. (2021). Being lean: how to shape digital transformation in the manufacturing sector. Journal of Manufacturing Technology Management, 32(9), 239-259. https://doi.org/10.1108/JMTM-12-2020-0467

[25] Bhadu, J., Bhamu, J., & Saraswat, P. (2023). An Analytic Hierarchy Process (AHP) Approach for Prioritizing the Industries 4.0 Technologies (I4. 0T). https://doi.org/10.5109/6792813

[26] Kaushik, M., & Mathur, B. (2014). Comparative study of K-means and hierarchical clustering techniques. Int. J. Softw. Hardw. Res. Eng, 2(6), 93-98.

[27] Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2023). Standardised metrics and methods for synthetic tabular data evaluation. Authorea Preprints. https://doi.org/10.36227/techrxiv.16610896.v1

[28] Riviére, C., & Marlair, G. (2010). The use of multiple correspondence analysis and hierarchical clustering to identify incident typologies pertaining to the biofuel industry. Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy, 4(1), 53-65. https://doi.org/10.1002/bbb.187

[29] Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. Educational Researcher, 38(5), 343-352. http://dx.doi.org/10.3102/0013189X09339056

[30] Durap, A. (2023). A comparative analysis of machine learning algorithms for predicting wave runup. Anthropocene Coasts, 6(1), 17. http://dx.doi.org/10.1007/s44218-023-00033-7

[31] Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. International Journal of Interactive Multimedia and Artificial Intelligence, 5, 26-31. http://dx.doi.org/10.9781/ijimai.2018.02.004